ABSTRACT

Title of dissertation: MODELING IMATINIB-TREATED CHRONIC MYELOGENOUS LEUKEMIA AND THE IMMUNE SYSTEM

Cara Disa Peters
Doctor of Philosophy, 2019

Dissertation directed by: Professor Doron Levy
Department of Mathematics

Chronic myelogenous leukemia can be considered as a chronic condition thanks to the development of tyrosine kinase inhibitors in the early 2000s. Most CML patients are able to manage the disease, but unending treatment can affect quality of life. The focus of much clinical research has thus transitioned to treatment cessation, where many clinical trials have demonstrated that treatment free remission is possible. While there are a lot of existing questions surrounding the criteria for cessation candidates, much evidence indicates the immune system plays a significant role.

Mathematical modeling provides a complementary component to clinical research. Existing models well-describe the dynamics of CML in the first phase of treatment where most patients experience a biphasic decline in the BCR-ABL ratio. The Clapp model is one of the first to incorporate the immune system and capture the often-seen oscillations in the BCR-ABL ratio that occur later in therapy. However, these models are far from capable of being used in a predictive manner and do

not fully capture the dynamics surrounding treatment cessation.

Based on clinical research demonstrating the importance of immune response, we hypothesize that a mathematical model of CML should include a more detailed description of the immune system. We therefore present a new model that is an extension of the Clapp model. The model is then fit to patient data and determined to be a good qualitative description of CML dynamics. With this model it can be shown that treatment free remission is possible. However, the model introduces new parameters that must be correctly identified in order for it to have predictive power.

We next consider the parameter identification problem. Since the dynamics of CML can be considered in two phases, the biphasic decline of and oscillations in the BCR-ABL ratio, we hypothesize that parameter values may differ over the course of treatment and look to identify which parameters are most variable by refitting the model to different windows of data. It is determined that parameters associated with immune response and regulation are most difficult to identify and could be key to selecting good treatment cessation candidates.

To increase the predictive power of our model, we consider data assimilation techniques which are successfully used in weather forecasting. The extended Kalman filter is used to assimilate CML patient data. Although we determine that the EKF is not the ideal technique for our model, it is shown that data assimilation methods in general hold promising value to the search for a predictive model of CML. In order to have the most success, new techniques should be considered, data should be collected more frequently, and immune assay data should be made available.

MODELING IMATINIB-TREATED CHRONIC MYELOGENOUS LEUKEMIA AND THE IMMUNE SYSTEM

by

Cara Disa Peters

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Doron Levy, Chair/Advisor
Professor Kayo Ide
Professor Konstantina Trivisa
Professor Pierre-Emmanuel Jabin
Professor Sergei Sukharev

Acknowledgments

I would like to thank my advisor Dr. Doron Levy first and foremost for introducing me to the exciting field of mathematical oncology. The combination of mathematics and medicine has provided me with a challenging and rewarding research experience and increased my appreciation for the power of mathematics. Dr. Levy has been open and supportive throughout my time in the AMSC program with regard to both research and personal interests. I am grateful to have had this opportunity to work with him.

I would also like to thank Dr. Franck Nicolini, Dr. Thomas Lepoutre, Dr. Geoff Clapp and Dr. Apollos Besse for their helpful introduction to both the clinical and mathematical sides of CML research. Their willingness to answer a variety of questions and engage in discussion was significant to the work presented in this dissertation.

Thank you to the members of my dissertation committee: Dr. Kayo Ide, Dr. Konstantina Trivisa, Dr. Pierre-Emmanuel Jabin and Dr. Sergei Sukharev. In particular, I would like to thank Kayo Ide for her instruction on data assimilation and how to best apply these methods to CML models and Konstantina Trivisa for her encouraging words and infectious love of mathematics.

There are many friends both inside and outside the department that I am grateful to have had throughout this process: Danielle Middlebrooks, Asia Wyatt, David Russell, Nick Julia, Shawna Bruell and many others. Special thanks to Matt Becker for a million conversations about math, research and life in general. Our

conversations throughout grad school and especially during the last few months have impacted me more than I can say. A special thanks also to Karla Lopez for being an amazing friend by always checking in on me and providing much needed pep talks.

Last but not least, I owe a huge thank you to my family, mom, dad, Britta and Mike, and my boyfriend Stephen who have always been supportive and encouraging. Throughout all of my frustrations and doubts, successes and failures, you have always believed I could accomplish my goals. I absolutely could not have done this without you. I am so thankful to have you all in my life.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

This dissertation aims to develop mathematical models and numerical methods that can be used to study the progression of Chronic Myeloid Leukemia (CML) under therapy, and ultimately be instrumental in evaluating and improving treatment protocols. We begin with an introduction to the biological properties and processes that serve as the foundation for this work.

1.1 Chronic Myeloid Leukemia

Chronic Myelogenous Leukemia (CML) is a type of blood cancer resulting in the overproduction of blood cells of the myeloid lineage. Approximately 15-20% of all leukemia cases in adults are CML [1, 2]. CML can be characterized by a genetic mutation in hematopoietic stem cells in which a translocation between chromosomes 9 and 22 occurs, fusing the BCR and ABL genes to form what is known as the Philadelphia (Ph) chromosome. This particular characteristic is detectable in more than 95% of all CML patients [3] and is used to confirm diagnosis. Without successful treatment, CML progress from chronic phase (CML-CP) through an accelerated phase (CML-AP) to a highly lethal blastic phase (CML-BP) within 3-5 years [4].

Fusion of the BCR-ABL gene is responsible for a variety of cancer-favorable changes, including changes in adhesion molecules, resistance to apoptosis, growth factor independence, and inhibition of DNA repair [5, 6]. Most notably, the BCR-ABL gene results in increased tyrosine kinase activity, which contributes to uncontrolled stem cell growth. The extent of damage caused by this oncogene makes it essential for development and maintenance of CML, and its effect on DNA repair is the likely cause for disease progression.

There are currently many types of treatments available to CML patients. Early treatment protocols relied on cytotoxic agents which are able to normalize white blood cell counts, but, in addition to harmful side effects, are usually incapable of preventing progression to blast phase. In the early 1980s, interferon-alpha (IFN) became the standard of treatment. Significant antitumor and immunomodulatory affects of IFN lead to a doubling of the median survival rate [4]. Despite these advances, few patients experienced a complete cytogenic response (CCR) and the adverse effects of IFN make it untenable for long treatment schedules.

Discovery of the Ph chromosome and specifically the BCR-ABL gene, contributed to an increase in reasearch of gene specific therapies and development of BCR-ABL tyrosine kinase inhibitors (TKIs). TKIs, such as imatinib, dasatinib, and nilotinib, specifically target Ph cells by occupying the binding pocket of the BCR-ABL protein to prevent phosphorylation of any substrates [5]. This controls the population of leukemic cells by inhibting cellular proliferation of mutated cells. Imatinib has proven to be a highly effective treatment by which approximately 80% of patients experience CCR [3] and is largely responsible for doubling the 5-year

survival rate [7]. Side effects of imatinib are much milder than previous therapies, but approximately 25% of patients develop resistance to it [4].

Second generation TKIs, like dasatinib and nilotinib, were developed to address imatinib resistance. Nilotinib is more than 30 fold more effective than imatinib at lysing BCR-ABL expressing cells and is effective against 32/33 imatinib-resistant point mutants [6]. Dasatinib is also more effective against BCR-ABL mutant types with high imatinib resistance. Despite these promising statistics, TKIs are not a cure for CML as they fail to eliminate leukemic stem cells (LSCs) which are primarily responsible for disease recurrence. Most CML patients will need to take the drug for life which can decrease their quality of life. Stem cell transplantation and donor transfusions remain the only potentially curative therapies available.

Clinicians classify response to treatment in three ways: hematological, cytogenetic and molecular. Hematological response is measured through blood cell counts, with a complete hematological response (CHR) characterized as a normalization of blood counts and spleen size. Cytogenetic response is determined by the decrease in Ph^+ metaphase cells in the blood. Patients with Ph^+ metaphase in 0-35% of cells are classified as having major cytogenetic response (MCR). An absence of Ph metaphase cells is considered to be a complete cytogenetic response (CCR). Patient response to therapy can also be assessed through measurement of BCR-ABL chimeric mRNA by real-time quantitative polymerase chain reaction (PCR). Quantitative PCR is the most sensitive assay to detect the presence of the BCR-ABL gene, detecting one cell in a background of $10^5 - 10^6$ normal cells [8]. Molecular response to therapy is assessed by this method. A 3-log reduction in the BCR-ABL ratio as measured by

real-time PCR is considered a major molecular response (MMR), while undetectable ratios are classified as complete molecular response (CMR). To account for variation in BCR-ABL ratios measured in different laboratories and therefore increase reliability of therapy protocols, an International Scale (IS) of measurement was adopted using method-specific conversion factors [9].

1.2 CML and the Immune System

In recent years much research has been conducted into the development of cancer immunotherapies, influenced by the cancer immune surveillance theory that the immune system has the ability to target tumor cells. The immune system is comprised of two components, innate and adaptive immunity, both of which are equally important to defending the body against foreign antigen. The innate immune system provides a primary rapid defense and plays a large role in activating the adaptive immune system, while adaptive immunity is more specific and therefore more effective at targeting various pathogens. Research and clinical studies have provided evidence both for and against the cancer immune surveillance theory [10], but the idea that the immune system can be harnessed or encouraged to destroy cancer cells remains. Evidence exists to suggest that this theory is relevant particularly to CML.

Much of the evidence of immune activity in CML is in the response of CML to immunomodulatory treatments such as donor lymphocyte tranfusions and stem cell therapy, which remains the only perceived cure for CML. Clinical trials with BCR-

ABL peptide vaccines have shown the possibility of provoking an anti-CML immune response [11]. TKI therapy also shows evidence of immune activity. A majority of patients with untreated CML have lymphocytes with clonal T cell receptor gene rearrangements. These clonal T cells continue to exist at low levels under imatinib therapy, while clonal $CD8^+$ cytotoxic cells and NK cells expand significantly under dasatinib therapy [11, 12]. In vitro studies have shown that TKIs can inhibit T cell activation and proliferation and have immunomodulatory effects on antigen presenting cells, but whether these inhibitory effects occur in patients is yet to be seen [6, 12].

Prior to the use of TKIs, interferon-alpha (IFNα) was the predominant treatment for CML patients. IFNα has many anti-tumor and immunomodulatory mechanisms, including inhibition of cell growth, induction of apoptosis, promotion of cycling of quiescent hematopoeitc stem cells, activation of immune effector cells, and increased expression of tumor associated antigen [11, 13]. Research has shown a significant increase in $CD8^+$ T cells and NK cells in patients undergoing IFNα therapy. The presence of CTLs specific for PR1 expressed by CML progenitors has also been noted in patients who receive IFNα [2]. For these reasons, IFNα is experiencing a revival of usage in CML treatment, specifically in combination with TKIs.

In addition, within the last decade the outcomes of various clinical trials on treatment cessation suggest that the immune system plays a role in combating CML. The Stop Imatinib (STIM) study [14] enrolled 100 CML patients undergoing imatinib treatment in CMR for at least two years as candidates for treatment cessation.

Of the 69 patients with at least 12 months of follow-up, 41% experienced persistent CMR at month 12. The remaining 42 patients relapsed, which was defined as positivity of BCR-ABL transcripts as measured by quantitative PCR with a BCR-ABL ratio of 10^{-5} or more in a second point of analysis. TWISTER [15] studied treatment cessation by following 40 patients that had been treated with imatinib and achieved undetectable minimal residual disease (UMRD) during treatment. Relapse was characterized as any single sample with BCR-ABL ratio greater than 0.1% or two consecutive samples with any detectable value. Of the 40 patients, 45% had not relapsed by time of analysis. Five of these patients had detectable BCR-ABL ratios on one or two occasions. Euro-Ski [16] enrolled 758 CP-CML patients using any TKI for at least 3 years with confirmed CMR for at least 1 year. Relapse was defined as loss of MMR, or BCR-ABL ratio greater than 0.1% on the International Scale. 61% and 50% of patients experience relapse-free survival at 6 months and 24 months respectively. All three of these cessation studies were relatively successful and demonstrated that treatment free remission (TFR) is possible. However, low levels of leukemic cells are often still detectable in these patients. Fluctuating BCR-ABL levels just below MMR but without loss of MMR, in common in many of these trials [11, 17] suggesting that some other mechanism, perhaps immune response, is responsible for keeping the disease under control and preventing a relapse.

1.3 Outline of Dissertation

In the past decade, there has been much interest in the use of mathematical models to gain further insight into the dynamics of CML genesis and to analyze the effects of treatment. These models are summarized by Clapp and Levy [18]. While many models focus on hematopoiesis, treatment and drug resistance, few models currently incorporate the role of the immune system in the growth and control of CML. Chapter 2 of this dissertation presents a review of a few of these models. Chapter 3 presents an extension of the Clapp *et al.* model, which first captured the oscillations observed in the leukemic loads of patients.

The remaining components of this dissertation focus on the use of numerical methods for improving the predictive ability of mathematical models of CML. These methods are applied to the extended model presented in Chapter 3. Chapter 4 incorporates the use of adaptive or adjustable parameter values. This will allow for better understanding of dynamic changes in immune response as well as better tools for the design of adaptive immunotherapy treatments and gradual treatment cessation. Chapter 5 explores methods of data assimilation to apply our model to treatment cessation data and forecast when relapse may occur. Chapter 6 provides the conclusion to the thesis.

Chapter 2: Mathematical Models of CML and the Immune System

2.1 Overview

The use of mathematical modeling and computational methods to study biology, medicine and the health sciences can have many benefits to how diseases are studied, diagnosed and treated. This chapter presents various models of CML and the immune system that have informed the work in this dissertation.

2.2 Models of CML

This section will provide a summary of existing mathematical models of CML that have served as a basis for the thesis work. We begin by looking at two models that factor in hematopoeisis but do not incorporate the immune system [19, 20]. These models, when looking at CML progression under imatinib therapy, show a biphasic exponential decline in the BCR-ABL ratio. While this behavior does correspond to phenomena seen in patient data, it does not completely capture the dynamics of the disease. Further, these models do not lend themselves to be good models of treatment cessation.

The following two models of CML presented here [21, 22], provide the main

foundation for the model introduced in Chapter 3. They also looks at CML on a cellular level, separating leukemia cells into different compartments based on their stage in the hematopoetic process, while also incorporating the immune system. These two models give evidence for an optimal range of leukemic cell concentration in which the immune response is most effective, a result that could greatly inform the use of immunotherapies in CML treatment. Using a simple representation of the immune response, the last model [22] describes the dynamics of CML beyond the reach of the other models [19, 20].

2.2.1 An Agent Based Model

Roeder *et al.* considered an agent based model of CML [19], biologically based on a cell differentiation process consisting of three stages: stem cells, precursors and mature cells. Stem cells are further categorized as either non-proliferating (A) or proliferating cells (Ω). Movement between these compartments corresponds with the development and maturation process of hematopoeitic cells (fig. 2.1). Each stem cell may be characterized by its cellular affinity, a quantity based on cell age and state. Non-proliferating or quiescent cells increase their affinity over time until maximal affinity is reached. Quiescent cells become proliferating, or cycling, stem cells with some probability ω determined by affinity and the total number of proliferating cells. Cycling stem cells proliferate by completing the 48 hour cell cycle. The cell cycle consists of four necessary phases for cell growth and division. These stages in order are G_1, S, G_2, and M. Roeder *et al.* considered quiescent cells to become

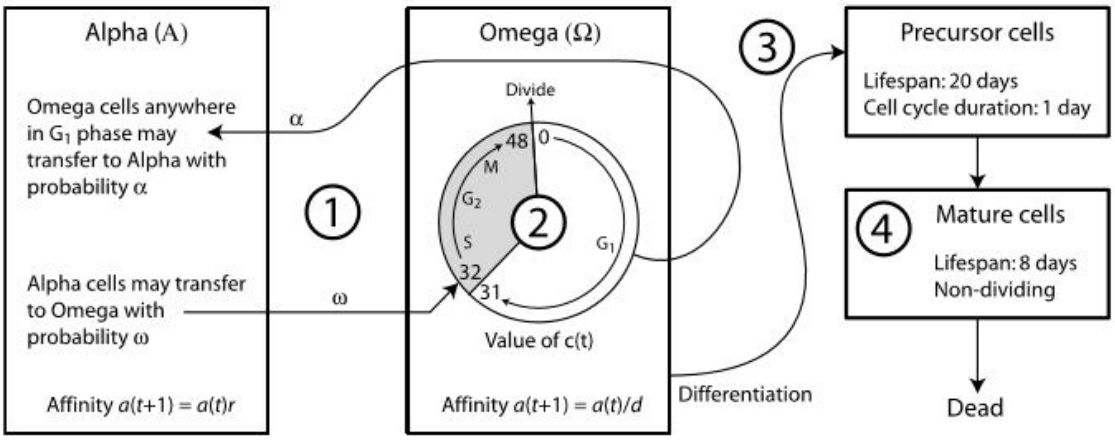


Figure 2.1: Diagram of hematopoetic cell differentiation process as described by Roeder *et al.* [19]. Cells are categorized as stem cells, precursors or mature cells. Stem cells are further delineated as quiescent (A) or cycling (Ω). Figure from [20]

proliferating at hour 32 of the cell cycle, the beginning of the S phase during which DNA synthesis occurs. At hour 48, the cell divides into two identical daughter cells that each begin the cycle in the G_1 growth phase. Transitions from Ω to A occur during the G_1 phase with probability α. Proliferating stem cells lose cellular affinity over time until minimal affinity is attained. The transition probabilities ω and α are determined by each cell's affinity and the total number of cycling or quiescent cells respectively:

$$\begin{aligned} \omega(\Omega(t), a(t)) &= \frac{a_{min}}{a(t)} f_\omega(\Omega(t)), \\ \alpha(A(t), a(t)) &= \frac{a(t)}{a_{max}} f_\alpha(A(t)). \end{aligned} \tag{2.1}$$

Stem cells with minimum affinity differentiate into precursor cells. Within this model, cells are considered to divide symmetrically once every 24 hours for 20 days,

at which point they differentiate into fully functional mature cells. Cells live in the mature stage for 8 days before dying.

The biology described here is a simplification of the cell maturation process and makes a few assumptions. First, the differentiation process has been reduced to three stages of maturation. Second, transition probabilities between stem cell compartments are assumed to be based on affinity, an internal quantity for each stem cell that varies in time within an interval $[a_{min}, a_{max}]$. Affinity is a notion whose existence was postulated by Roeder [19] and is not directly associated with any known biological mechanism specific to the hematopoietic system. Furthermore, the time spent in each stage is deterministic. It is assumed that these lifespans are known and fixed. Lastly, it is assumed that when a cell undergoes mitosis and divides, it does so symmetrically. This means that each daughter cell is identical and of the same type of cell as the parent, i.e. stem cells divide into two stem cells, precursors divide into two precursors.

An agent based model (ABM) provides a natural way of describing the hematopoietic system as it can mathematically represent the underlying biological system as a single-cell stochastic process. It retains information about individual cell characteristics and interactions with other cells. The ABM presented by Roeder *et al.* [19] defines a set of rules to govern the development of hematopoietic stem cells, as described above. These rules are applied to each stem cell simultaneously in discrete time steps that begin by computing the total number of both A and Ω stem cells. These values are used to govern the behavior of each stem cell in the model. The transition probabilities ω and α for cells in A and Ω respectively are calculated by

equation (2.1). A stem cell is considered to transition from A to Ω or vice versa if the transition probability is greater than a randomly generated number. Cells that remain in A increase affinity by a factor r known as the regeneration factor while cells in Ω decrease affinity by a differentiation factor $1/d$. Unlike the stem cells in this model, the behavior of differentiated cells is not stochastic. Therefore, these cells do not need to be treated on an individual basis.

The rules described above can be used to simulate a population of healthy hematopoietic cells. For a long enough simulation, healthy stem cells reach a steady state profile in which the majority of cells are quiescent stem cells having maximal affinity (fig. 2.2a). The Roeder *et al.* ABM is applied to CML and imatinib therapy by simulating three non-interacting cell populations: healthy cells (Ph^-), leukemic cells (Ph^+) and imatinib-affected cells ($Ph^{+/A}$).

Since Ph^+ cells proliferate uncontrollably, the transition rates between A and Ω differ from those of Ph^- cells. This difference is incorporated with distinct parameter values for the sigmoidal transition functions $f_{\alpha/\omega}$ found in the transition probabilities (2.1). A majority of simulations show Ph^+ mature cells overtaking the Ph^- mature cell population approximately 5 years after the introduction of leukemia (fig. 2.2b). Due to the stochasticity of the ABM, not all simulations result in this outcome (fig. 2.3a). This model is able to capture delayed CML growth as well as early eradication of the Ph^+ population.

Simulation of imatinib treatment initiates once the proportion of differentiated Ph^+ cells reaches more than 99.5% of the total cell population. The model incorporates the effects of imatinib in two ways, reflecting the two ways in which imatinib

may affect leukemic cells. Cycling Ph^+ stem cells become imatinib-affected with probability r_{inh} and both Ph^+ and $Ph^{+/A}$ cells undergo apoptosis with probability r_{deg}. This is implemented as a stochastic decision at the beginning of each time step of the ABM.

The affects of treatment are measured using a ratio of leukemic cells to healthy cells given by the BCR-ABL ratio:

$$\text{BCR-ABL Ratio} = \frac{\#\text{ mature Ph}^+\text{ cells}}{\#\text{ mature Ph}^+\text{ cells} + 2 \cdot (\#\text{ mature Ph}^-\text{ cells})}$$

Simulations of treatment depict a biphasic exponential decline of this ratio over a treatment span of 400 days (figs. 2.2c and 2.3b), which is consistent with the dynamics seen in patient data. This model can be quite computationally expensive as the complexity is proportional to the number of cells being simulated. In order to achieve relatively efficient runtime at initial implementation in 2006 [19], cell numbers had to be scaled down to approximately $\frac{1}{10}$ of realistic numbers, thereby simulating approximately 10^5 stem cells. More efficient implementations of this algorithm were proposed by Kim *et al.* in [20, 21] and are described in detail below.

2.2.2 A System of Difference Equations

The limitations of the ABM were considered and improved by Kim *et al.* [20]. Rather than simulating each stem cell individually, this model groups cells by their common characteristics and describes their behavior as a system of discretized difference equations. In order to devise the difference equations, the state space must

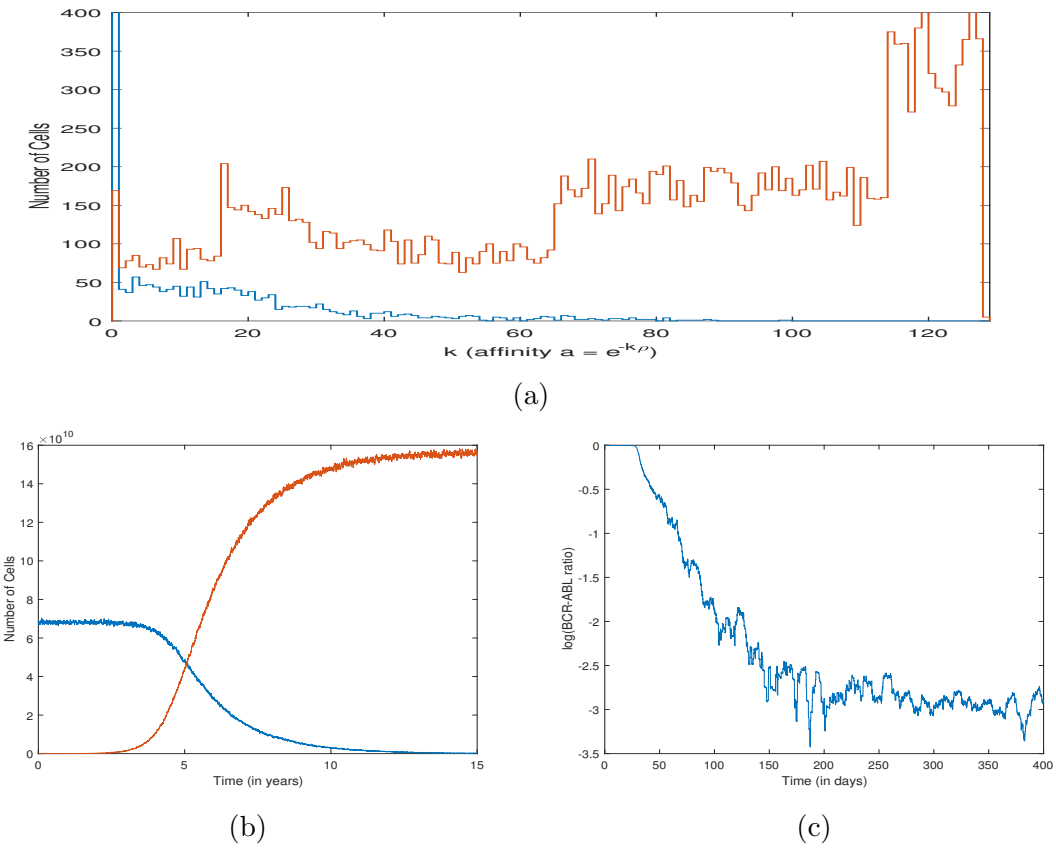
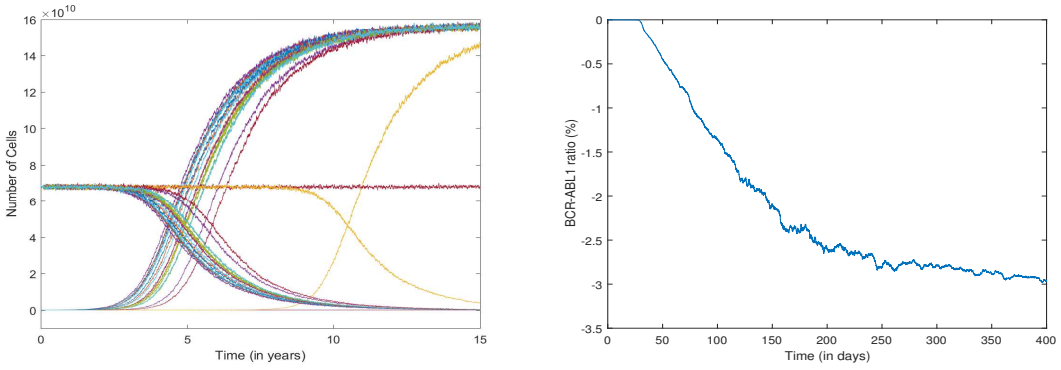


Figure 2.2: Example simulation of the ABM. (a) Steady state profile of stem cells plotted as number of quiescent (orange) and cycling (blue) stem cells versus affinity level. (b) Mature Ph^+ (orange) and Ph^- (blue) cells during CML geneis. (c) BCR-ABL ratio plotted over a 400 day treatment period.



(a) (b)

Figure 2.3: Simulation of CML for 20 different runs of the ABM over a 15 year period. (a) Mature Ph^+ and Ph^- cells populations. Corresponding runs are plotted in the same color. Almost all simulations reach a 99% BCR-ABL threshold within the 15 year period. One simulation predicts slower CML development, while the last predicts the extinction of the leukemic population after roughly 100 days of simulation. (b) Average BCR-ABL ratio of 18 ABM simulations plotted over a 400 day simulation of treatment. Only runs that attained a 99% BCR-ABL threshold were used to calculate the average.

be discretized. Affinity is discretized by setting $a(t) = e^{-k\rho}$ where $\rho = \log d \approx 0.0488$ and $0 \leq k \leq 127$ is an integer. The affinity of each cell can therefore be represented discretely by the value of k where $\log a(t) = -k\rho$. The stem cell populations are modeled by the following difference equations with k representing cell affinity and c representing position in the cell cycle.

$$A_k(t+1) = \begin{cases} (A_0(t) - B_0(t)) + (A_1(t) - B_1(t)) + (A_2(t) - B_2(t)), & k = 0 \\ (A_{k+2}(t) - B_{k+2}(t)) + \sum_{c=0}^{31} \Psi_{k,c}(t), & k = 1, ..., 125 \\ \sum_{c=0}^{31} \Psi_{k,c}(t), & k = 126, 127 \end{cases} \tag{2.2}$$

$$\Omega_k(t+1) = \begin{cases} B_0(t), & k = 0, c = 32 \\ 2\Omega_{k-1,48}(t), & k > 0, c = 0 \\ \Omega_{k-1,c-1}(t) - \Psi_{k-1,c-1}(t), & k > 0, c = 1, ..., 31 \\ (\Omega_{k-1,31}(t) - \Psi_{k-1,31}(t)) + B_k(t), & k > 0, c = 32 \\ \Omega_{k-1,c-1}(t), & k > 0, c = 33, ..., 48 \\ 0, & otherwise \end{cases} \tag{2.3}$$

Transitions between the A_k and $\Omega_{k,c}$ compartments are determined by binomial random variables B_k and $\Psi_{k,c}$ which have the following distributions:

$$B_k(t)\ Bin(A_k(t), \omega(\Omega(t), e^{-k\rho})$$

$$\Psi_{k,c}(t)\ Bin(\Omega_{k,c}(t), \alpha(A(t), e^{-k\rho})$$

Here $\Omega(t) = \sum_{k,c} \Omega_{k,c}$ and $A(t) = \sum_k A_k(t)$ denote the total number of cycling and quiescent cells respectively. The transition probabilities ω and α are as given in (2.1).

The differentiated cells are represented in a similar fashion. The equations for precursors are denoted $P_j(t)$ where $j = 0, ..., 479$ is the number of hours a cell has spent as a precursor, up to 20 days. Similarly, mature cells are denoted by $M_j(t)$ where $j = 0, ..., 191$.

$$P_j(t+1) = \begin{cases} \sum_{c=0}^{48} \Omega_{127,c}(t) - \sum_{c=0}^{31} \Psi_{127,c}(t), & j = 0 \\ 2P_{j-1}(t), & j = 24, 48, 72, ..., 456 \\ P_{j-1}(t), & otherwise \end{cases} \tag{2.4}$$

These equations directly reflect the rules of cell differentiation as described in Section 2.2.1. For example, the first line of (2.4) represents cycling stem cells that have attained minimum affinity and differentiate into precursors. Precursors divide every 24 hours producing two identical daughter cells, as represented by line 2 of (2.4). Line 3 denotes an increase in age of the precursors, which is necessary to track the time spent as a precursor before the cell matures.

As in the ABM (2.2.1), three non-interacting cell populations are simulated to mathematically model clinically observed phenomena. The equations as given above ((2.2)-(2.4)) mathematically describe the growth and development of Ph^- cells. Simulation of this model produces a quiescent-dominated steady state stem cell profile (fig. 2.4a), similar to the ABM. Altering the parameter values in (2.3)

allows the model to replicate uncontrolled cell growth and therefore model Ph^+ and $Ph^{+/A}$ cells (fig. 2.4b).

To incorporate the effects of imatinib therapy, equation (2.3) is altered for Ph^+ and $Ph^{+/A}$ cells. Let $\Omega^{+/I}(t)$ $Bin(\Omega^+_{k,c}(t), r_{inh})$ be the number of proliferating Ph^+ stem cells infected by imatinib as time t. Let $\Omega^{+/D}(t)$ $Bin(\Omega^+_{k,c}(t), r_{deg})$ be the number of proliferating Ph^+ stem cells undergoing apoptosis at time t. Then, the right hand side of (2.3) is replaced by $\Omega^{+/R}_{k,c}(t)$ where $\Omega^{+/R}_{k,c}(t) = \Omega^+_{k,c}(t) - \Omega^{+/I}_{k,c}(t) - \Omega^{+/D}_{k,c}(t)$ is the number of cycling Ph^+ stem cells remaining unaffected at the next time step. Likewise, for $Ph^{+/A}$ cells, the right hand side of (2.3) is replaced by $\Omega^{+/A,R}_{k,c}(t) = \Omega^{+/A}_{k,c}(t) + \Omega^{+/I}_{k,c}(t) - \Omega^{+/A,D}_{k,c}(t)$. The overall structure of the equations remains unchanged. This model also captures the biphasic decline typical of the BCR-ABL ratio typical under TKI therapy (fig. 2.4c).

The computational efficiency of this model lends itself well to exploring the various effects of imatinib therapy. Simulation of the Kim *et al.* difference equation model ((2.2)-(2.4)) during therapy shows an immediate and rapid decline in the number of Ph^+ cells (fig. 2.5a). However, this model does not predict a complete eradication of CML. In an average run of this model, the Ph^+ population reaches a minimum value of approximately 5.6×10^6 cells after 160 days of treatment. Extending the treatment period beyond 400 days result in a continuously declining BCR-ABL ratio (fig. 2.5b). Since a steady state is never reached in any simulation, the model suggests that eradication of CML is possible if duration of therapy is long enough. This result does not agree with the true behavior of CML reflected in available patient data. Since treatment cessation is an important question in

current clinical research of CML, it is important to look at what this model can say about post-treatment dynamics. A 400 day treatment period significantly reduces the leukemic cell population. When treatment simulation is stopped (all model transitions occur according to the original transition rules), the Ph^+ population is able to recover and outgrow the Ph^- population approximately 5 years post-treatment (fig. 2.5c). By 10 years post-treatment in all simulations of the model, the BCR-ABL ratio has reached the 99% barrier once more, suggesting a complete relapse of CML (fig. 2.5d). The Kim *et al.* difference equation model is unable to simulate TFR. This therefore suggests that this model, and similarly the Roeder ABM, is lacking in its ability to capture the full range of dynamics of this disease.

2.2.3 The Kim Model

Kim *et al.* also formulated a model [21] to provide an alternative explanation for the long-term remission experienced by most CML patients undergoing imatinib therapy, by incorporating the effects of immune response on CML. The inclusion of immune response was inspired by experimental research that suggested an anti-leukemia immune response by CD4+ and CD8+ T cells. The model [21] is

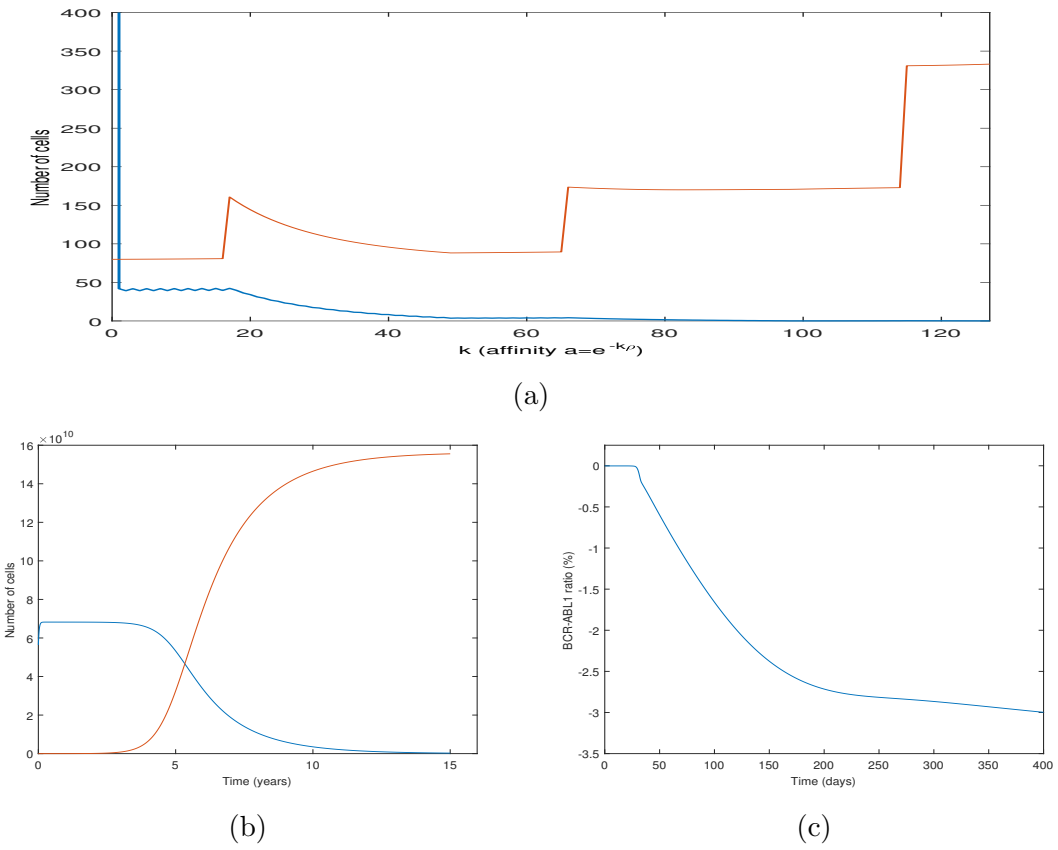


Figure 2.4: Example simulation of the Kim *et al.* difference equation model [20]. (a) Steady state profile of stem cells plotted as number of quiescent (orange) and cycling (blue) stem cells versus affinity level. (b) Mature Ph^+ (orange) and Ph^- (blue) cells during CML geneis. (c) BCR-ABL ratio plotted over a 400 day treatment period.

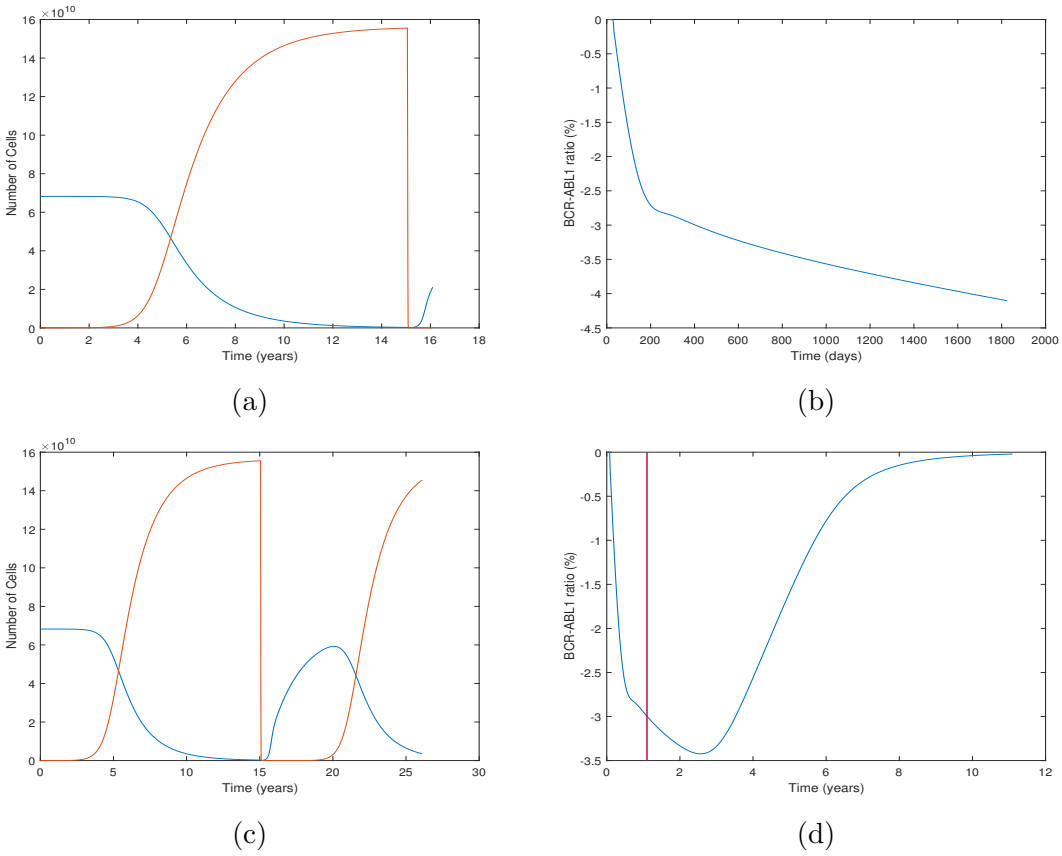


Figure 2.5: A look at the effects of imatinib therapy by simulation of the Kim *et al.* difference equation model [20]. (a) Mature Ph^+ (orange) and Ph^- (blue) cell populations over a 15 year simulation of CML genesis and 400 day treatment period. (b) Log of the BCR-ABL ratio plotted over 5 year treatment period. (c) Mature Ph^+ (orange) and Ph^- (blue) cell populations over 15 year CML genesis phase, 400 day treatment period and 10 year post treatment span. (d) Log of the BCR-ABL ratio (blue) corresponding to the mature cell populations displayed in (c). The red line indicated treatment cessation.

formulated as a system of DDEs:

$$\dot{y_0} = [r_y - d_0]y_0 - q_C p(C,T)y_0, \tag{2.5a}$$

$$\dot{y_1} = ay_0 - d_1 y_1 - q_C p(C,T)y_1, \tag{2.5b}$$

$$\dot{y_2} = by_1 - d_2 y_2 - q_C p(C,T)y_2, \tag{2.5c}$$

$$\dot{y_3} = cy_2 - d_3 y_3 - q_C p(C,T)y_3, \tag{2.5d}$$

$$\dot{T} = s_T - d_T T - p(C,T)C + 2^n p(C_{n\tau}, T_{n\tau}) q_T C_{n\tau}. \tag{2.5e}$$

Leukemic cells (y_i) are compartmentalized into stem cells (y_0), progenitors (y_1), differentiated cells (y_2), and terminally differentiated cells (y_3), to describe the hematopoietic differentiation process. Stem cells grow at a rate r. Progenitors, differentiated cells and terminally differentiation cells differentiate and expand at rates a, b, and c respectively. These differentiation rates are reduced to simulate imatinib therapy. Natural death rates for each cell compartment are given by d_i.

Anti-leukemia response is modeled by the final term $q_C p(C,T)y_i = q_C p_0 e^{-c_n C} kTy_i$ in each of the leukemic cell equations where C denotes the total concentration of all leukemic cells and T is the total concentration of anti-leukemia T cells. Each leukemic cell population interacts with T cells at a rate kTy_i. T cells engage the leukemia cell with probability p_0 and kills it with probability q_C. The exponential component of these terms reflect suppression of immune response by leukemia cells.

The T cell population is established from a constant supply term s_T and dies off at a rate d_T. T cells are stimulated to amplify upon successful interaction with

leukemic cells. Stimulated T cells exit the system, reentering after completing n divisions. The average duration of one division is given by τ. $C_{n\tau}$ and $T_{n\tau}$ are the concentrations of leukemic and T cells after a delay of $n\tau$ time units.

With this model 2.5, Kim *et al.* described CML dynamics and long-term remission through incorporation of anti-leukemia immune response. Furthermore, the structure of immune-leukemia interaction terms allowed them to study the balance between the two cell types and determine an optimal load zone in which the anti-leukemia T cell response is most effect. This optimal load zone is a precursor to the immune window hypothesized in the Clapp model [22].

2.2.4 The Clapp Model

Clapp *et al.* noticed that while much of their patient data showed the typical biphasic exponential decline described by the models in sections 2.2.1 and 2.2.2 and other mathematical models, the data also showed fluctuations in BCR-ABL ratio after this period that existing models were unable to capture. Representative graph for two patients are shown in Figure 2.6. Clapp hypothesized that these oscillations were due to additional mechanisms of the autologous immune system. They thus

constructed an ODE model [22] of CML and the immune system:

$$\dot{y_0} = b_1 y_1 - a_0 y_0 - \frac{\mu y_0 z}{1 + \epsilon y_3^2}, \tag{2.6a}$$

$$\dot{y_1} = a_0 y_0 - b_1 y_1 + r y_1 (1 - \frac{y_1}{K}) - d_1 y_1 - \frac{\mu y_1 z}{1 + \epsilon y_3^2}, \tag{2.6b}$$

$$\dot{y_2} = \frac{a_1}{inh_1} y_1 - d_2 y_2 - \frac{\mu y_2 z}{1 + \epsilon y_3^2}, \tag{2.6c}$$

$$\dot{y_3} = \frac{a_2}{inh_2} y_2 - d_3 y_3 - \frac{\mu y_3 z}{1 + \epsilon y_3^2}, \tag{2.6d}$$

$$\dot{z} = s_z - d_z z + \frac{\alpha y_3 z}{1 + \epsilon y_3^2}. \tag{2.6e}$$

The model consists of four leukemic cell populations, quiescent stem cells (y_0), cycling stem cells (y_1), progenitors (y_2), mature cells (y_3), and a single immune cell compartment (z). The maturation process of leukemic cells generally follows the assumptions previously described in Sections 2.2.1 and 2.2.2. Stem cells are either quiescent or cycling and can transition between the two types until they differentiate into progenitors. Cycling stem cells are assumed to follow logistic growth. Progenitor cells further differentiate into mature cells. Each leukemic cell type is assumed to have a natural death rate and also dies due to interaction with immune cells. The effect of TKI therapy is incorporated by reducing the differentiation rate and expansion factors for progenitors and mature cells. The immune system is simplistically represented as a single compartment, supplied by a constant source s_z with a natural death rate d_z. Amplification of immune cells is directly stimulated by leukemic cell presence. Clapp *et al.* introduce the concept of an immune window to describe the window of leukemic load in which the immune system is most

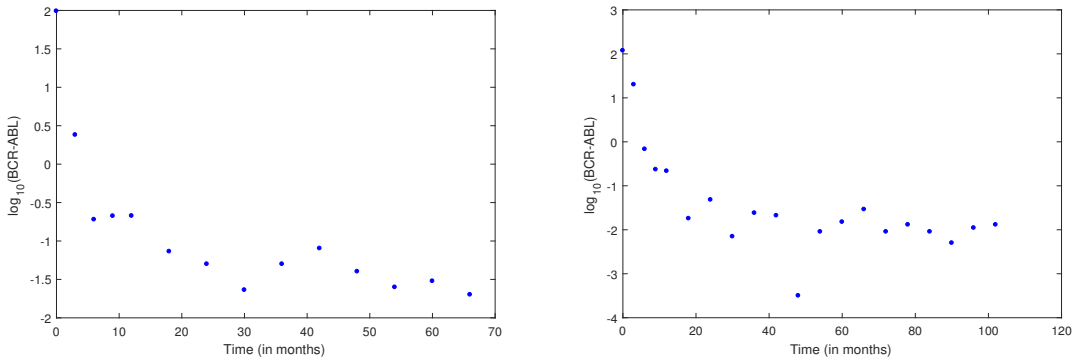


Figure 2.6: Representative patient data for two CML patients. The log of BCR-ABL ratio during imatinib therapy is plotted over time in months. An initial biphasic decline can be seen, followed by oscillations in the ratio.

active and effective. Large leukemic populations reduce amplification of immune cells while small leukemic populations affect the ability of immune cells to destroy leukemic cells. It is assumed that immune cells affect all types of leukemic cells in the same manner.

With this mathematical model, Clapp *et al.* were able to not only capture the same dynamics of previous models, notably the biphasic decline, but they were also able to demonstrate that the oscillations in the BCR-ABL ratio observed in a majority of patient data, may be the result of an individualized response to the drug and a varying degree of the immune response.

2.3 Modeling the Dynamics of the Immune System

In the last 10-15 years, the study of immunology through use of mathematical models has grown drastically. The nature of these models varies greatly from

the scale of the model (molecular, cellular, tissue, multi-scale), to the mathematics behind it (ODEs, PDEs, Monte-Carlo simulations), and the specific area of immunology being studied (T cell receptors, tumor-immune interaction, etc.). Many of these variations and example models are laid out in various reviews [23–25]. In this section, we will briefly discuss a delay differential equation (DDE) model of adaptive regulation of the immune response on a cellular level and various extensions of this model. This model will serve as the motivation for the CML model in Chapter 3.

2.3.1 A Model of Adaptive Regulation

Kim *et al.* proposed a mathematical model of primary immune response incorporating contraction of the immune system through regulatory T cells [26]. The model is formulated as a system of DDEs describing the interaction of five cell types: naive antigen presenting cells A_0 (APCs), mature APCs A_1, naive T cells K_0, mature T cells K, and regulatory T cells R. The model is as follows:

$$\dot{A}_0 = s_A - d_0 A_0(t) - a(t)A_0(t), \tag{2.7a}$$

$$\dot{A}_1 = a(t)A_0(t) - d_1 A_1(t), \tag{2.7b}$$

$$\dot{K}_0 = s_K - \delta_0 K_0(t) - kA_1(t)K_0(t), \tag{2.7c}$$

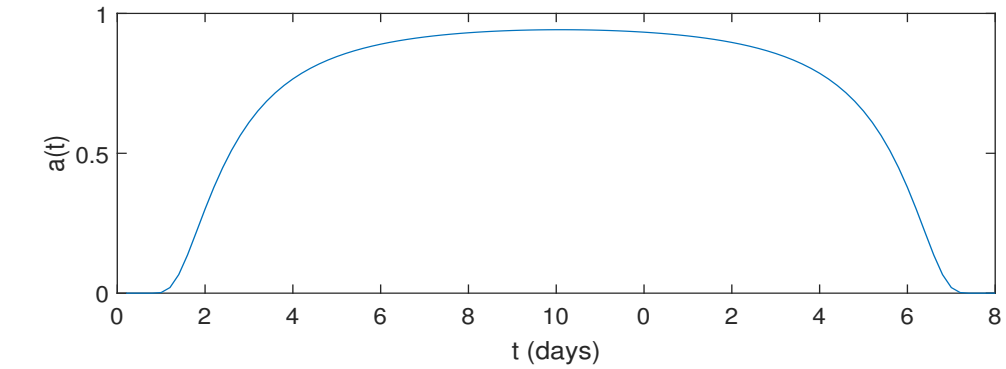
$$\begin{aligned}\dot{K} = {} & 2^m kA_1(t-\sigma)K_0(t-\sigma) - kA_1(t)K(t) + 2kA_1(t-\rho)K(t-\rho) \\ & - (\delta_1 + r)K(t) - kR(t)K(t),\end{aligned} \tag{2.7d}$$

$$\dot{R} = rK(t) - \delta_1 R(t). \tag{2.7e}$$

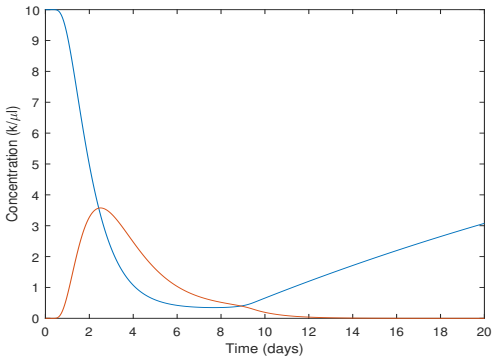
Naive APCs interact with antigen at a rate $a(t)$, causing them to mature and transport the antigen to the lymph node where immature T cells are located. The presence of APCs in the lymph initiates a minimal developmental program. After m divisions, T cells are considered fully developed and further proliferate due to interaction with mature APCs. Mature T cells differentiate into regulatory T cells at a rate r. Regulatory cells suppress mature T cells. Naive APCs and T cells are supplied by constant source terms s_A and s_K respectively and all cells type are assumed to experience a natural death.

This model does not distinguish between CD4+ and CD8+ T cells in order to simplify the model and thereby focus on the feedback between effector cells and regulatory cells in general. Kim *et al.* note that their formulation of differentiation of effector cells into regulatory cells is based on the notion of an antigen-independent T cell program [26]. That is, a primary immune response is initially stimulated by the presence of antigen but T cells are then able to self-regulate independently of further stimulation from APCs. Although this model represents an oversimplification of the biological processes that occur in a primary immune response, it is able to describe the basic dynamics (fig. 2.7), providing a useful tool for studying immune regulation.

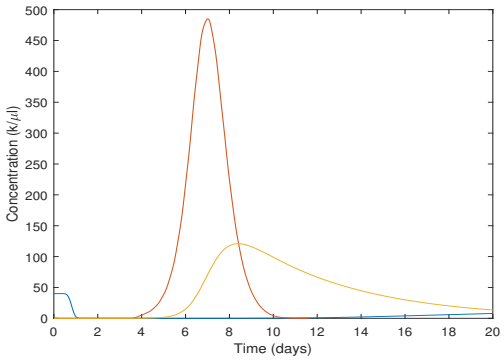
Kim *et al.* also established an extension of this model [26], incorporating CD4+ T



(a)



(b)



(c)

Figure 2.7: (a) Graph of $a(t)$, the rate at which antigen interact with and stimulate immature APCs. (b) Dynamics of immature and mature APCs over a 20 day period. (c) Dynamics of naive effector, mature effector and regulatory T cells. To view the dynamics of all three populations on one figure the concentrations of naive cells and regulatory cells are multiplied by a factor of 1000 and 100 respectively.

cells and stimulation of T cell expansion by cytokine:

$$\dot{A}_0 = s_A - d_0 A_0(t) - a(t)A_0(t), \tag{2.8a}$$

$$\dot{A}_1 = a(t)A_0(t) - d_1 A_1(t), \tag{2.8b}$$

$$\dot{H}_0 = s_H - \delta_0 H_0(t) - kA_1(t)H_0(t), \tag{2.8c}$$

$$\begin{aligned}\dot{H} = {} & 2^{m_1} kA_1(t-\sigma_1)H_0(t-\sigma_1) - kA_1(t)H(t) + 2kA_1(t-\rho_1)H(t-\rho_1) \\ & - (\delta_H + r)H(t) - kR(t)H(t),\end{aligned} \tag{2.8d}$$

$$\dot{K}_0 = s_K - \delta_0 K_0(t) - kA_1(t)K_0(t), \tag{2.8e}$$

$$\begin{aligned}\dot{K} = {} & 2^{m_2} kA_1(t-\sigma_2)K_0(t-\sigma_2) - kP(t)K(t) + 2kP(t-\rho_2)K(t-\rho_2) \\ & - \delta_K K(t) - kR(t)K(t),\end{aligned} \tag{2.8f}$$

$$\dot{P} = r_1 H(t) + r_2 K(t) - \delta_P P(t) - kP(t)K(t) - kP(t)R(t), \tag{2.8g}$$

$$\dot{R} = rH(t) - kP(t)R(T) + 2kP(t-\rho_1)R(t-\rho_1) - \delta_H R(t). \tag{2.8h}$$

Here H_0, H and P represent the concentrations of naive CD4+ (helper) T cells, mature CD4+ T cells and positive growth signal, respectively. The dynamics of naive and mature APCs remain the same as in Equation (2.7). Mature APCs activate both types of effector cells, initiating their minimal developmental programs. Upon completing a specific number of divisions, further interaction with mature APCs cause CD4+ effector cells to continue dividing, while CD8+ effector cells proliferate by consuming positive growth signal. Positive growth signal is secreted by both types of effector cells. Effector CD4+ cells differentiate into regulatory T cells. Proliferation of regulatory cells is stimulated by positive growth signal and regulatory

cells suppress both CD4+ and CD8+ effector cells. This more detailed model better represents the underlying biology, yet generally yields the same dynamics as the original model.

Chapter 3: A Model of CML and the Immune System under TKI Treatment

3.1 Introduction

The purpose of this chapter is to introduce and study an extension of the Clapp *et al.* model (2.6) that was introduced in Chapter Section 2.2.4. Clapp *et al.* [22] constructed their model of CML and the autologous immune system in order to address the inability of previous models [19, 20] to capture certain key characteristics of CML patient data. They hypothesized that the immune system was in some way responsible for the development of fluctuations in the BCR-ABL ratio seen in much of the available data.

While model (2.6) produces a relatively good fit to the data and supports the hypothesis that the immune system plays a role in the dynamics of CML under IM therapy, it is based on an overly simplified representation of immune response. The goal of this chapter is to extend the model of [22] by replacing the single immune compartment with a more accurate representation of immune response. Including immune cell subtypes should improve the accuracy of the fit of the model to the clinical data. It will also provide the necessary infrastructure for incorporating

detailed clinical data regarding the immune dynamics of individual patients once immunological assays are made available.

Our proposed model of CML with expanded immune components is presented in Section 3.2. In Section 3.3 we discuss parameter estimation and demonstrate simulations of the model. A discussion of the results, including a comparison to the Clapp *et al.* model, can be found in Section 3.4.

3.2 An Improved Model of the Immune Response to CML

As in the Clapp *et al.* model, we let y_0, y_1, y_2 and y_3 represent concentrations of quiescent leukemic stem cells, cycling leukemic stem cells, progenitor leukemic cells and mature leukemic cells, respectively. We replace the immune cell compartment z with three new compartments T_0, T_1 and R that represent the concentrations of naive T cells, mature T cells and regulatory T cells, respectively.

The dynamics of the leukemic cell populations remain largely unchanged from the base model. Stem cells transition between quiescent and cycling states, and experience growth while cycling. Cycling stem cells may differentiate into progenitor cells, which may further differentiate into mature cells. Cycling stem cells, progenitors and mature cells all have a natural death rate. Leukemic cells are also killed through interaction with mature effector T cells. From a constant source, naive T cells are activated and develop into mature T cells based on interaction with leukemic cells. Further interaction with leukemic cells prompts mature T cells to continue dividing. Mature T cells differentiate into regulatory cells which can

suppress mature T cell growth.

For simplicity, both CD4+ and CD8+ T cells are grouped together into a single effector T cell compartment, and other types of effector cells (NKs, etc.) are excluded. This simplification does not allow us to represent the specific mechanisms by which regulatory cells are produced, which is not a concern of this work. We assume that effector cells that differentiate into regulatory cells do not proliferate further. Inclusion of the regulatory cells in this model allows us to handle the contraction of the immune response independently of the CML cell populations. These simplifications and assumptions are motivated by the adaptive immune regulation model of Kim *et al.* [26]. This allows us to effectively simulate an immune response while leaving open the possibility of further expansion in later work through inclusion of other immune cell types, biologically accurate dynamics, and incorporation of immunotherapies.

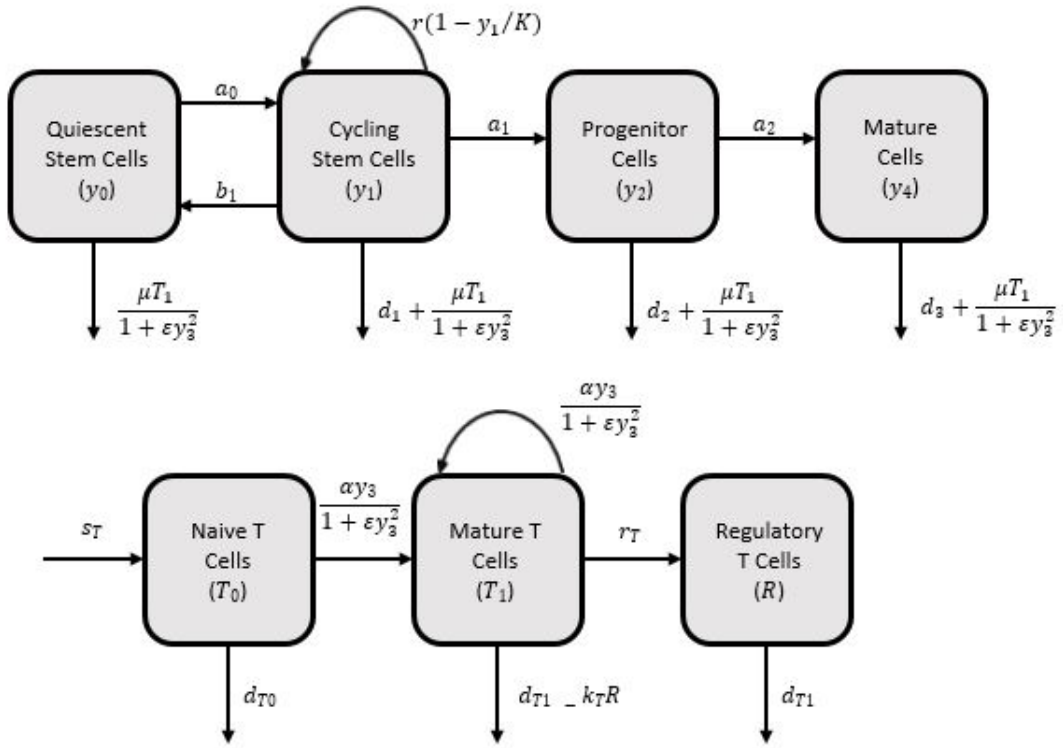


Figure 3.1: Diagram of mathematical model of CML with expanded immune compartments. Leukemic cells are divided into four subpopulations: quiescent stem cells (y_0),cycling stem cells (y_1), progenitor cells (y_2) and mature cells (y_3). Immune cells are divisded into three subtypes: naive T cells (T_0), mature T cells (T_1) and regulatory T cells (R). Leukemic stem cells transition between quiescence and cycling, undergoing divison and duplication in the cycling compartment. Cycling stem cells can differentiate into progentior cells, which can then differentiate into mature cells. Leukemic cells die at a natural death rate or due to interaction with mature T cells. We assume a constant supply rate of naive T cells. Naive cells are activated and develop into mature T cells do to the presence of leukemic cells. Mature T cells continue to proliferate in response to further interaction with leukemic cells and differentiate into regulatory T cells at rate r. Regulatory T cells suppress mature T cells and all immune cells die with natural death rates d_{Ti}.

The model is structured as the following system of ODEs:

$$\dot{y_0} = b_1 y_1 - a_0 y_0 - \frac{\mu y_0 T_1}{1 + \epsilon y_3^2}, \tag{3.1a}$$

$$\dot{y_1} = a_0 y_0 - b_1 y_1 + r y_1 (1 - \frac{y_1}{K}) - d_1 y_1 - \frac{\mu y_1 T_1}{1 + \epsilon y_3^2}, \tag{3.1b}$$

$$\dot{y_2} = \frac{a_1}{inh_1} y_1 - d_2 y_2 - \frac{\mu y_2 T_1}{1 + \epsilon y_3^2}, \tag{3.1c}$$

$$\dot{y_3} = \frac{a_2}{inh_2} y_2 - d_3 y_3 - \frac{\mu y_3 T_1}{1 + \epsilon y_3^2}, \tag{3.1d}$$

$$\dot{T_0} = s_T - d_{T_0} T_0 - \frac{\alpha y_3 T_0}{1 + \epsilon y_3^2}, \tag{3.1e}$$

$$\dot{T_1} = \frac{\alpha y_3 T_0}{1 + \epsilon y_3^2} + \alpha y_3 T_1 - (d_{T_1} + r_T) T_1 - k_T R T_1, \tag{3.1f}$$

$$\dot{R} = r_T T_1 - d_{T_1} R. \tag{3.1g}$$

Equation (3.1a) describes the dynamics of quiescent leukemic stem cells. These cells enter the cell cycle at a rate b_1 and return to quiescence at a rate a_0. Quiescent stem cells undergo apoptosis after interaction with mature T cells at a rate μ. Equation (3.1b) refers to cycling stem cells. The first two terms are similar to those in (3.1a), describing the transition between the two stem cell compartments. We use logistic growth to describe the proliferation of cycling stem cells, with growth rate r and carrying capacity K. Cycling stem cells have a natural death rate of d_1 and are also killed by mature T cells with maximal effective kill rate μ. The first term of equation (3.1c) describes the differentiation of cycling stem cells into progenitors. The parameter value a_1 incorporates both the differentiation rate and the amplification that occurs due to proliferation. Progenitor cells die at a natural death rate d_2 as well as by interaction with mature T cells. The fourth equation

describes the dynamics of mature leukemic cells and is similar to equation (3.1d). Here, the differentiation rate is a_2 and the death rate is d_3.

The final three equations of the model govern the dynamics of the immune cell compartments. Equation (3.1e) includes terms for a constant source s_T and natural death rate d_{T0} of naive T cells. Naive T cells are activated through interaction with leukemic cells at a rate α. The first term of (3.1f) represents the development of naive T cells into mature cells, while the second term refers to proliferation of mature T cells stimulated by leukemic cells. Mature T cells die with natural death rate d_{T1} and differentiate into regulatory cells at a rate r_T. The last term refers to the suppression of mature T cells by regulatory cells at a rate k_T. Finally, equation (3.1g) describes the dynamics of regulatory cells. The first term gives the rate at which mature T cells differentiate into regulatory cells and the second describes the natural death of regulatory cells. We assume that this death rate is the same as that of mature T cells.

As in the Clapp model [22], this model incorporates immunosupression in two ways. First, mature leukemic cells inhibit the killing potential of mature T cells. This can be seen in the final term of (3.1a)-(3.1d) where μ is divided by a factor of $1 + \epsilon y_3^2$. Second, mature leukemic cells diminish the activation of naive effector T cells and the proliferation of mature effector T cells. This is incorporated by decreasing the expansion rate α by the same factor of $1 + \epsilon y_3^2$ in (3.1e) and (3.1f). Structuring immunosupression in this way restricts the effectiveness of effector T cells to an immune window $[y_{min}, y_{max}]$. Large leukemic loads ($y_3 > y_{max}$) result in an ineffective immune response, while leukemic populations that are too small

($y_3 < y_{min}$) do not stimulate enough of a response. Since the total leukemic cell population is predominantly mature cells, we consider only the y_3 concentration when incorporating immunosuppression.

Imatinib targets CML in two ways: reducing the proliferation rate and triggering apoptosis of leukemic cells. This model incorporates the former by reducing the proliferation rates a_1 and a_2 of progenitor and mature leukemic cells by factors of inh_1 and inh_2 respectively. There is little consensus on the effects of imatinib on leukemic stem cells with some research suggesting that CML stem cells may be partially or completely resistant to this therapy [27]. Therefore, we do not incorporate an effect of imatinib on the stem cell compartments of our model.

3.3 Results

3.3.1 Data and Numerical Methods

We fit our model to patient data made available by our collaborator Dr. Franck Nicolini (Centre Léon Bérard). The data set consists of BCR-ABL ratios of 104 CML patients that were monitored during imatinib therapy in the Centre Hospitalier Lyon Sud. Patients were treated with first-line imatinib 400 mg daily. BCR-ABL ratios were measured in the same laboratory by quantitative RT-PCR. Measurements were taken at diagnosis, every 3 months for the first year of therapy, and every 6 months thereafter. Each ratio was measured twice and the average was reported. Of the 104 patients in this study, 33 patients changed TKIs for safety reasons and 14 patients experienced a progression in disease, with a few patients falling into both categories.

We excluded these patients and fit our model to the remaining 65 patients who responded well to imatinib. This is the same data set used by Clapp *et al.* and therefore allows for comparison of the two models.

Most of the parameters in our model are considered to be universal; their values are estimated based on background biological knowledge and are fixed for all patients (Table 3.1). The remaining eight parameters relate to either the effect of imatinb therapy (inh_1, inh_2) or immune response (μ, d_{T_1}, α, ϵ, k_T, r_t). Since the immune system varies between individuals [28] and the effectiveness of imatinib differs between patients, we consider these parameters to be patient specific and are therefore determined by fitting the model to measured patient BCR-ABL ratios. Latin hypercube sampling is used to generate sample parameter sets. For each patient, a single parameter set is selected by minimizing the squared distance between the log of the simulated ratio and the log of the measured ratio. This parameter set is then used as an initial condition in a nonlinear minimization routine to find the optimal patient-specific parameter set.

Rather than fit the immune expansion rate α and immunosupression strength constantϵ directly, we fit the immune window $[y_{min}, y_{max}]$ and compute $\epsilon = 1/(y_{min}y_{max})$ and $\alpha = (y_{min} + y_{max})\epsilon d_{T1}$. The sampling range for y_{max} can be found by determining the y_3 concentration at the second phase of decline in the BCR-ABL ratio, since we hypothesize that this decline occurs when the leukemic load is pushed into the immune window. Similarly, we look at the center and magnitude of the oscillations in the BCR-ABL ratio to determine an initial sampling range for y_{min}. For most patients in the available data set, these ranges are $[10^4, 10^6]$ for y_{max} and $[10^2, 10^4]$

for y_{min}.

We use Matlab's *ode45* differential equation solver to run simulations of our model. From the results of the solver, we compute a simulated BCR-ABL ratio with the following formula:

$$BCR - ABL\,ratio = 100\beta\frac{y_3}{2x + y_3} \tag{3.2}$$

This is a ratio of the number of BCR-ABL genes to normal ABL genes in the blood. Here, in (3.2), y_3 refers to the mature leukemic cell population, as in (3.1), x represents the healthy cell population, and β is a factor that accoutns for differences in mRNA expression between the mutated and control gene. Each cell has two copies of the ABL gene, one in each copy of chromosome nine. Only one copy is mutated when the Philadelphia chromosome is formed [29].

3.3.2 Simulations

We fit our model (3.1) to the available data for each patient using the process described in Section 3.3.1. The simulated BCR-ABL ratio for 12 CML patients undergoing Imatinib therapy is plotted on a logarithmic scale in Figures 3.2 and 3.3. Patient-specific parameters for these simulations are found in Table 3.2. From these figures, it is noticeable that this extended model can reproduce patient dynamics. Notably, the model captures the biphasic decline in BCR-ABL ratio typically seen during the first year of Imatinib treatment. The steep first phase of this decline occurs on average over the first 6 months. The second phase is more gradual,

Parameter	Value	Description	Source
a_0	0.0027	Transition rate to cycling stem cell	Estimated
b_1	0.0247	Transition rate to quiescent stem cell	Estimated
r	0.08	Logistic growth rate of cycling stem cell	Estimated, [21]
K	4.2872	Carrying capacity of cycling stem cell	Estimated
a_1	899.9820	Differentiation rate and expansion factor of progenitor cells	Estimated
a_2	24.0005	Differentiation rate and expansion factor of mature cells	Estimated
d_1	0.00225	Natural death rate of cycling stem cells	[21]
d_2	0.006	Natural death rate of progenitor cells	[21]
d_3	0.0375	Natural death rate of mature cells	[21]
d_{T0}	0.03	Natural death rate of naive T cells	[26]
s_T	$120 * d_{T0}$	Naive T cell source term	Estimated
$y_0(0)$	37.5	Initial concentration of quiescent stem cells	Estimated, [30]
$y_1(0)$	4.1667	Initial concentration of cycling stem cells	Estimated, [30]
$y_2(0)$	$1.6667 \cdot 10^4$	Initial concentration of progenitor cells	[30]
$y_3(0)$	$1.5 \cdot 10^8$	Initial concentration of mature cells	[31]
$T_0(0)$	120	Initial concentration of naive T cells	Estimated, [31]
β	3	BCR-ABL ratio adjustment factor	Estimated based on patient data

Table 3.1: Universal parameter values and descriptions.

sometimes appearing more as a plateau than a reduction (e.g. fig. 3.2a and fig. 3.3c). For some patients, the initial reduction in cancer load presents more as a triphasic decline, with the third phase being more gradual than the second. This can be seen in Figures 3.2b, 3.2e and 3.3e. The first phase again lasts approximately 6 months while the second phase varies in length from approximately 18 months (figs. 3.2e and 3.3e) to 34 months (fig. 3.2b). This initial deline in BCR-ABL ratio, either biphasic or triphasic, can be attributed to the effect of imatinib.

Of the 65 patients in the available data set, 15 have monotonically decreasing BCR-ABL ratios. The remaining patients experience fluctuations in their measured ratios, which the extended model is also able to capture. The effect of therapy pushes the cancer load into the immune window. This can occur within the first year (fig. 3.2a), two years (fig. 3.3f), or much later (fig. 3.2b) into therapy. At this

point, an immune response is initiated and effector T cells begin to kill CML cells. The combination of imatinib and immune response reduce the leukemic load to a concentration that lies below the immune window, ending the immune response and allowing a recurrence of CML. As the concentration of CML cells reenters the immune window, a second immune response begins and again the cancer load is pushed to a level below the immune window. This behavior continues, resulting in oscillations in the simulated BCR-ABL ratio. The nature of these oscillations vary depending on the patient. Period length can be quite short (fig. 3.2f) or much longer (fig. 3.3f). Amplitude of the oscillations often rapidly decreases (figs. 3.2a, 3.2d and 3.2f) until the simulated ratio approaches an equilibrium at y_{min}. Occasionally this steady state takes longer to approach, as in Figure 3.2c where visible oscillations are still occurring at year 10 of the simulation.

The involvement of the immune system is more clearly depicted in Figure 3.4. Fluctuations in the mature leukemic cell concentration lead to fluctuations in the concentration of mature effector T cells (fig. 3.4a). As the leukemic load initially drops into the immune window, an increase in mature effector T cells occurs corresponding to an initial immune response. The mature effector population decreases once the leukemic load is below the immune window and no longer stimulating activation and amplification of the effector cells. The largest spike in effector cells occurs during the initial immune response. As the concentration of effector cells oscillates, the amplitude of the oscillations decays over time. This occurs in our simulation due to the structure of our model and the lower concentration of leukemic cells, but it also makes sense biologically as a secondary immune response requires less effector

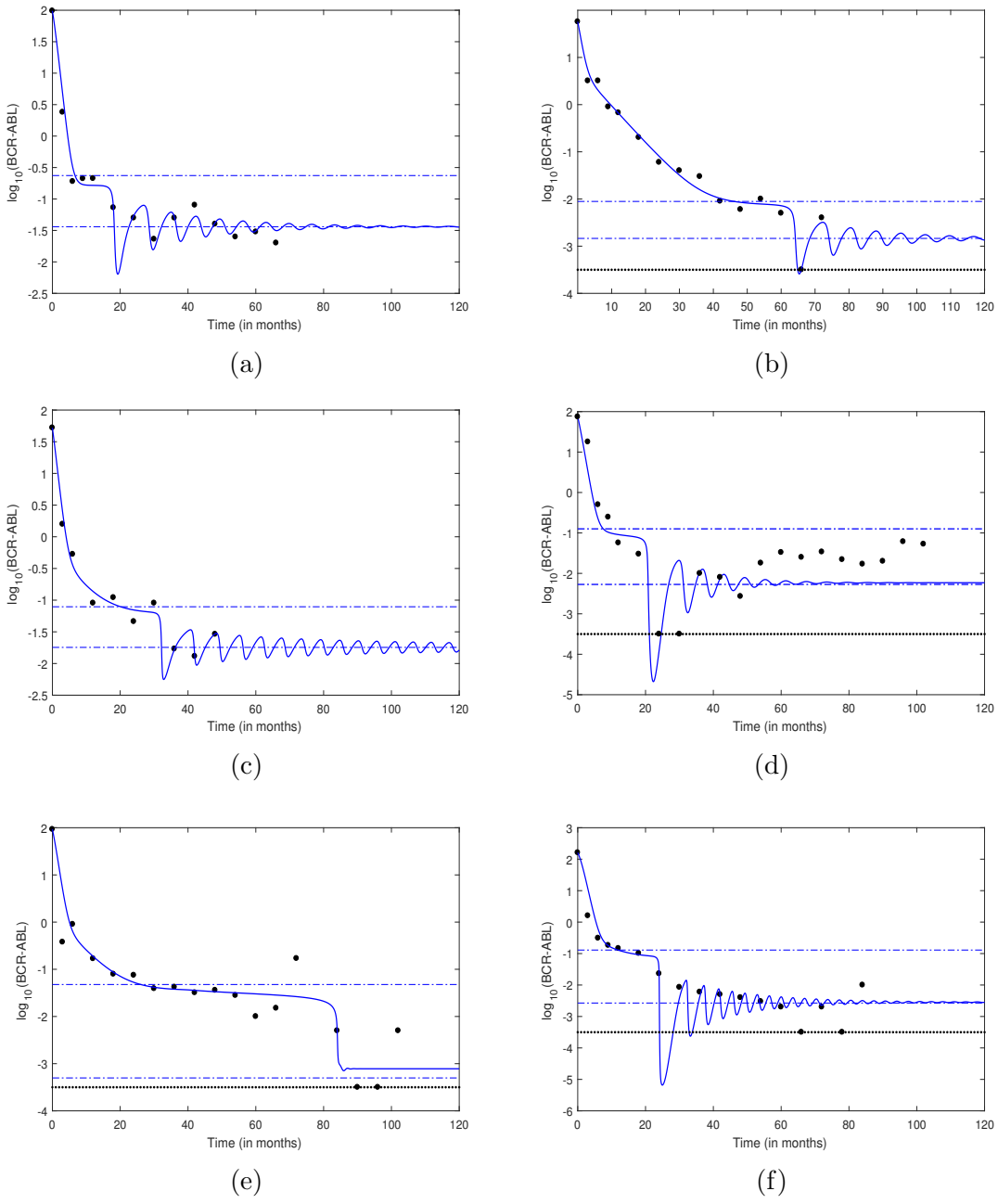


Figure 3.2: Fits of the the extended model (3.1a)-(3.1g) to patient data for six patients. The log of the BCR-ABL ratio is plotted over time. Dots represent patient data. Blue horizontal lines indicate the immune window. Black horizontal lines indicate the minimal level detectable by RT-PCR. Points are this line correspond to those points at which the disease was not observed.

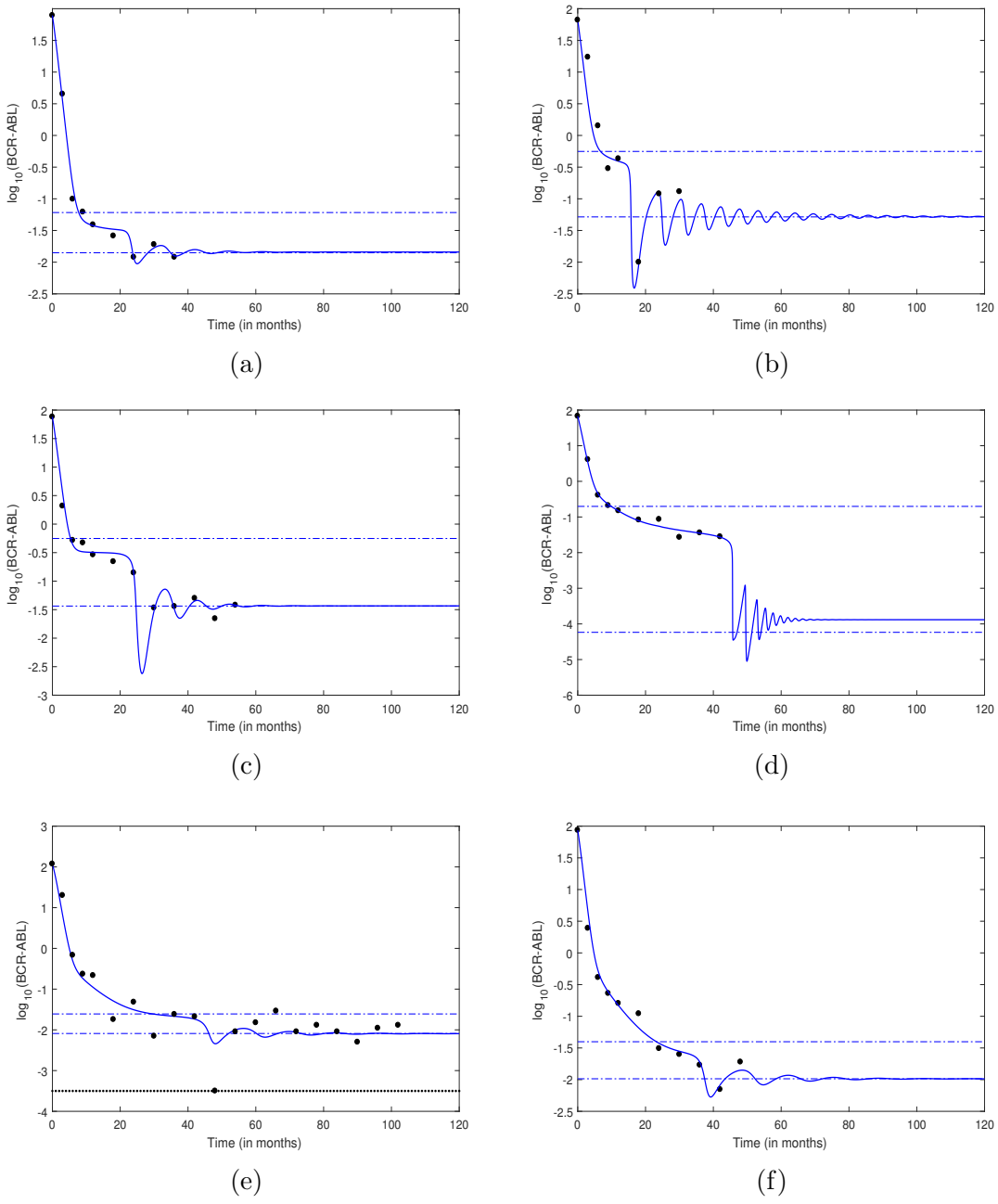


Figure 3.3: Fits of the the extended model (3.1a)-(3.1g) to patient data for six additional patients. The log of the BCR-ABL ratio is plotted over time. Dots represent patient data. Blue horizontal lines indicate the immune window. Black horizontal lines indicate the minimal level detectable by RT-PCR. Points are this line correspond to those points at which the disease was not observed.

cells to reduce the amount of antigen. The dynamics of each immune cell type are shown in Figure 3.4b. A reduction in naive effector T cells occurs at the beginning of an immune response, followed by a large spike in mature effector cells. Spikes in regulatory cell concentration occur towards the end of an immune response.

Since the model is quite complex and the number of parameters is large, sensitivity analysis of the parameters was done computationally rather than analytically. Figure 3.5 shows the quality of fit of the extended model as a function of four different pairs of patient-specific parameters for a single representative patient. The model is most sensitive to (inh_1, inh_2) and (y_{min}, y_{max}). This makes sense as inh_1 and inh_2 are responsible for the initial decline in leukemic load and y_{min} andy_{max} indicate when the immune system kicks on. There is also a strong correlation between fit and the parameters governing regulation of the immune system (r_T, k_T, d_{T1}). Although the range of acceptable values for these parameters seems to be much larger with pairs of values producing either very good or very poor fits. The maximal kill rate of effector cells (μ) seems to be less important; almost no correlation is seen between quality of fit and this parameter value.

3.4 Discussion

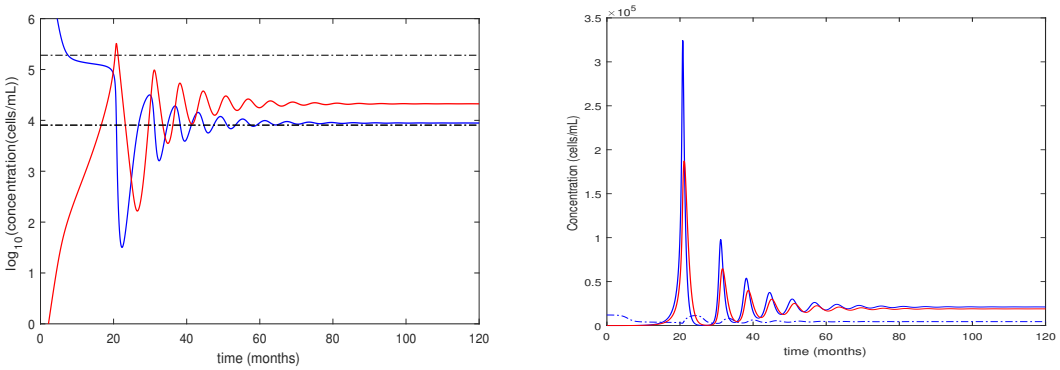
Since their introduction in the early 2000s, TKI therapies such as imatinib have greatly improved the outcomes for most CML patients. Despite improved survival rates and their impressive ability to reduce leukemic load, TKIs are not a cure for CML and the majority of patients rely daily on this drug to manage the

Patient	inh_1	inh_2	d_{T1}	μ	y_{min}	y_{max}	k_T	r_T
1	1.015	867.543	0.124	$7.045 \cdot 10^{-9}$	3.747e4	$2.441 \cdot 10^{5}$	$2.725 \cdot 10^{-5}$	$2.087 \cdot 10^{-7}$
2	644.798	14.424	0.097	$2.261 \cdot 10^{-7}$	$3.099 \cdot 10^{3}$	$1.884 \cdot 10^{4}$	0.003	$1.650 \cdot 10^{-8}$
3	10.231	100.858	0.374	$1.055 \cdot 10^{-9}$	$4.240 \cdot 10^{4}$	$1.855 \cdot 10^{5}$	1.417	$2.245 \cdot 10^{-11}$
4	2.280	558.193	0.047	$9.489e \cdot 10^{-7}$	$8.059 \cdot 10^{3}$	$1.908 \cdot 10^{5}$	$1.137 \cdot 10^{-7}$	0.001
5	32.815	113.547	0.143	$8.063 \cdot 10^{-10}$	$5.578 \cdot 10^{2}$	$5.331 \cdot 10^{4}$	$2.440 \cdot 10^{-7}$	0.001
6	6.340	743.093	0.114	$1.761 \cdot 10^{-9}$	$1.141 \cdot 10^{3}$	$5.515 \cdot 10^{4}$	$1.328 \cdot 10^{-8}$	0.003
7	2.661	1289.170	0.173	$6.826 \cdot 10^{-13}$	$1.993 \cdot 10^{4}$	$8.597 \cdot 10^{4}$	3.505	$2.374 \cdot 10^{-14}$
8	2.891	93.428	0.132	$3.032 \cdot 10^{-8}$	$9.151 \cdot 10^{4}$	$9.899 \cdot 10^{5}$	2.092	$1.597 \cdot 10^{-10}$
9	1.144	282.370	0.027	$3.325 \cdot 10^{-8}$	$5.399 \cdot 10^{4}$	$8.282 \cdot 10^{5}$	0.220	$4.626 \cdot 10^{-11}$
10	19.122	104.663	0.053	$3.200 \cdot 10^{-17}$	$1.003 \cdot 10^{2}$	$3.424 \cdot 10^{5}$	$1.628 \cdot 10^{-11}$	$1.409 \cdot 10^{-7}$
11	32.634	282.314	0.088	$7.787 \cdot 10^{-7}$	$6.249 \cdot 10^{3}$	$1.887 \cdot 10^{4}$	1.411	$2.092cdot10^{-11}$
12	40.701	126.263	0.093	$9.594 \cdot 10^{-8}$	$1.276 \cdot 10^{4}$	$4.906 \cdot 10^{4}$	0.139	$1.007 \cdot 10^{-10}$

Table 3.2: Patient parameter values used for the simulations in Figure 3.2 and Figure 3.3. Patients 1-6 correspond to Figures 3.2a to 3.2f. Patients 7-12 correspond to Figures 3.3a to 3.3f. The bounds of the immune window are fit to the data rather than the parameters α and ϵ. These parameters can be computed from the immune window as $\epsilon = 1/(y_{min}y_{max})$ and $\alpha = (y_{min} + y_{max})\epsilon d_{T1}$.

	inh_1	inh_2	d_{T_1}	μ	y_{min}	y_{max}	k_T	r_T
Min	0.398	10.548	0.016	$3.200 \cdot 10^{-17}$	100.309	$1.668 \cdot 10^{4}$	$1.628 \cdot 10^{-11}$	$2.374 \cdot 10^{-14}$
Mean	41.319	$1.497 \cdot 10^{3}$	0.114	$1.380 \cdot 10^{-7}$	$1.695 \cdot 10^{4}$	$3.291 \cdot 10^{5}$	4.029	$3.842 \cdot 10^{-4}$
Max	644.798	$9.496 \cdot 10^{3}$	0.374	$9.489 \cdot 10^{-7}$	$9.396 \cdot 10^{4}$	$1.968 \cdot 10^{6}$	84.5288	0.0042
STD	130.663	$2.511 \cdot 10^{3}$	0.078	$2.710 \cdot 10^{-7}$	$2.748 \cdot 10^{4}$	$4.788 \cdot 10^{5}$	17.178	$9.916 \cdot 10^{-4}$

Table 3.3: The minimum, mean, maximum and standard deviation are computed from the patient-specific parameter values of 25 of the patients in the data set. These 25 patients did not change TKI or dose during therapy, experienced no disease progression, and have at least 6 IS data points.



(a) (b)

Figure 3.4: A look at the immune compartments for a representative patient (Patient 4 in Figure 3.2d and Table 3.2). (a) depicts the concentration (cells/mL of blood) of mature CML cells (blue) and mature effector T cells (red) on a logarithmic scale over time. The immune window is plotted in black. (b) shows the dynamics of naive effector T cells (blue, dashed), mature effector T cells (blue solid), regulatory cells (red). Concentration (cells/mL of blood) is plotted over time. Note: the naive cell concentration is scaled by a factor of 100 and the regulatory cell concentration is scaled by a factor of 10 to show all three populations on the same graph.

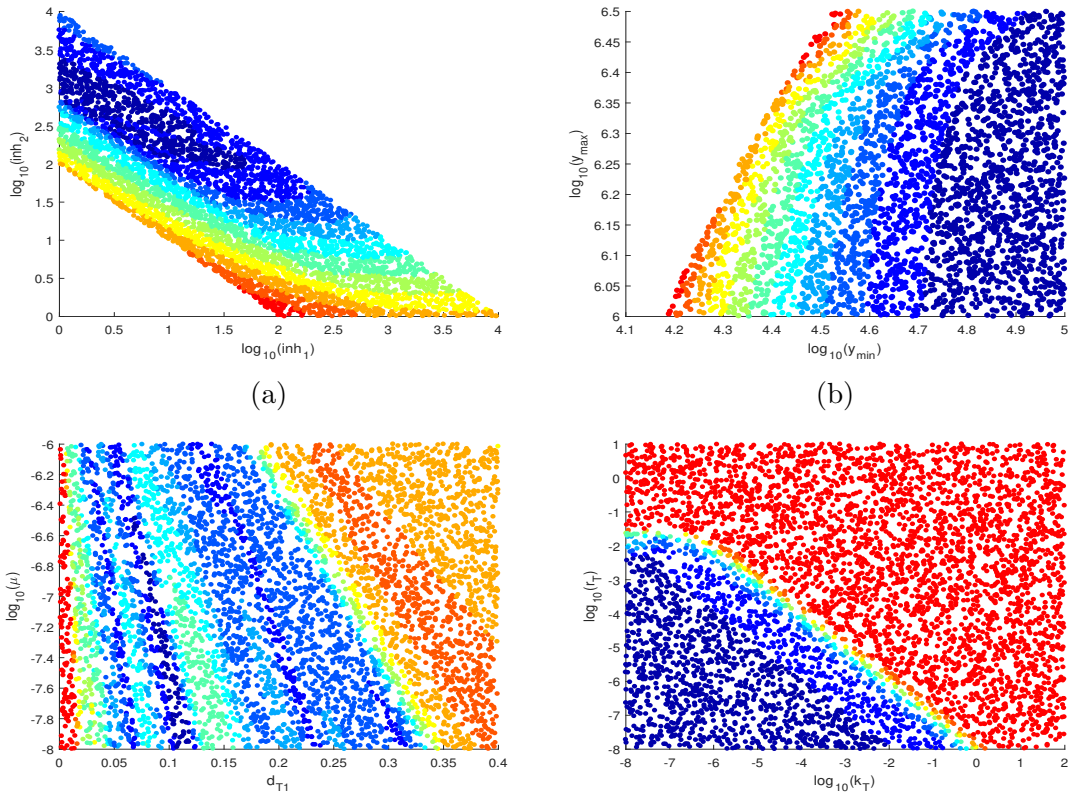


Figure 3.5: Sensitivity analysis of patient-specific parameters for a representative patient (Patient 1 in Figure 3.2a and Table 3.2). For each plot, the two specified patient-specific parameter values are varied while all other values are as stated in Table 3.1 and Table 3.2. The 2-norm error of the model (3.1) is plotted as a color. Better fits (lower error) are dark blue. Worse fits (higher error) are dark red. (a) $log(inh_1)$ vs. $log(inh_2)$ (b) $log(y_{min})$ vs. $log(y_{max})$ (c) d_{T1} vs. $log(\mu)$ (d) $log(k_T)$ vs. $log(r_T)$

disease. Recent studies such as STIM, TWISTER and EuroSki have shown the possibility of successful treatment cessation with over 40% of enrolled participants achieving TFR, but TFR does not equate to cancer free status. These patients often exhibit signs of CML presence, although at very low levels, suggesting other mechanisms are involved in maintaing TFR. These studies as well as other clinical evidence suggest that the immune system may be capable of combatting a rise in leukemic cell populations, and has aided a push to incorporate immunomodulating treatments like IFNα and further explore treatment cessation. The application of mathematical models can be a valuable tool in studying the connection between CML, immune response and successful treatment cessation.

Previous mathematical models, a few of which are described in Chapter 2, have provided valid descriptions of CML dynamics under TKI therapy, capturing a biphasic exponential decline in BCR-ABL ratio shortly after the onset of treatment. The Clapp *et al.* model [22] further explained characteristics of patient data by incorporating the immune system as a single compartment of their ODE model. Their hypothesis of a narrow window of leukemic cell concentration in which the immune system is both active and effective, was shown to produce oscillatory dynamics in the BCR-ABL ratio similar to those exhibited in available patient data. This assumption was further validated by Fassoni *et al.* in their comparison of models of the interaction of CML and immune effector cells [32]. Of the 20 different sub-models considered, they found that only those that utilized an immune window for either immune response to leukemic cells or immune recruitment were most plausible in that these models were capable of simulating three relevant steady states of disease,

remission and cure. The concept of an immune window is a solid foundation from which to work, but more is needed to be known about the interaction between the immune system and CML.

This chapter has presented a new mathematical model describing the dynamics of CML under imatinib therapy while incorporating immune response. The model retains the leukemic cell differentiation process of the Clapp model from stem cell to fully developed mature cell. Since it has been suggested that LSCs are a possible cause of relapse for CML patients, their inclusion in this model is vital. The single immune cell compartment of the Clapp model is replaced by three new compartments that describe both effector cell development and adaptive regulation through the inclusion of naive T cells and regulatory T cells. Expansion of the immune system compartments from previous models allows for a more biologically accurate description of immune response, and is a natural first step in model expansion.

While this extended model does appear to capture the dynamics of the disease, comparison of the model only to the measured BCR-ABL ratio provides little evidence to suggest it produces the true dynamics of CML under TKI therapy. Simulations of the Clapp model also give a generally valid depiction of the dynamics. Figures 3.6 and 3.7 provide a comparison of the two models for twelve representative patients. The two models sometimes predict similar or identical representations of CML under therapy (figs. 3.6a and 3.6b) and are close fits to the data. However, for other patients, the two models result in noticeably different outcomes. The expanded model can predict increased oscillatory BCR-ABL behavior (figs. 3.6c, 3.6f and 3.7b) or smoother approaches to a steady state along the immune window

minimum (figs. 3.7e and 3.7f), in comparison to the Clapp model. In either case, the 2-norm error between the model-simulated BCR-ABL ratio and the clinically-measured ratio is always small, and quite close to the errors produced by the Clapp model (table 3.4). This suggests that more information, perhaps more data or better computational methods, is needed for fitting these models to data in order to determine exactly what kind of dynamics are taking place and therefore increase the predictive value of these models.

Patient	1	2	3	4	5	6	7	8	9	10	11	12
Error - Clapp	0.560	0.556	0.466	1.991	1.635	1.493	0.355	0.773	0.397	0.174	1.651	0.435
Error - Extension	0.530	0.530	0.412	2.382	1.803	1.796	0.302	0.816	0.385	0.298	1.662	0.411

Table 3.4: Comparison of 2-norm errors for the Clapp *et al.* model (2.6) and our extended model (3.1).

An advantage of this extended model is its ability to provide insight into the patient immune profile, allowing the user to explore the effect of specific immune mechanisms on the behavior of CML under TKI therapy. Comparing patient-specific immune parameters confirms variation exists between the immune systems of different patients and suggests that the immune system can or does play a larger role in the success of TKI therapy for various patients. To further explore this hypothesis, we considered the removal of the regulatory T cell compartment. Regulatory T cells are responsible for suppressing immune response by decreasing the number of effector cells when antigen levels diminish. In this model, this occurs as the CML population falls below the immune window. Reduced levels of leukemic cells do not result in an eventual eradication of the disease. Instead, suppression of the immune

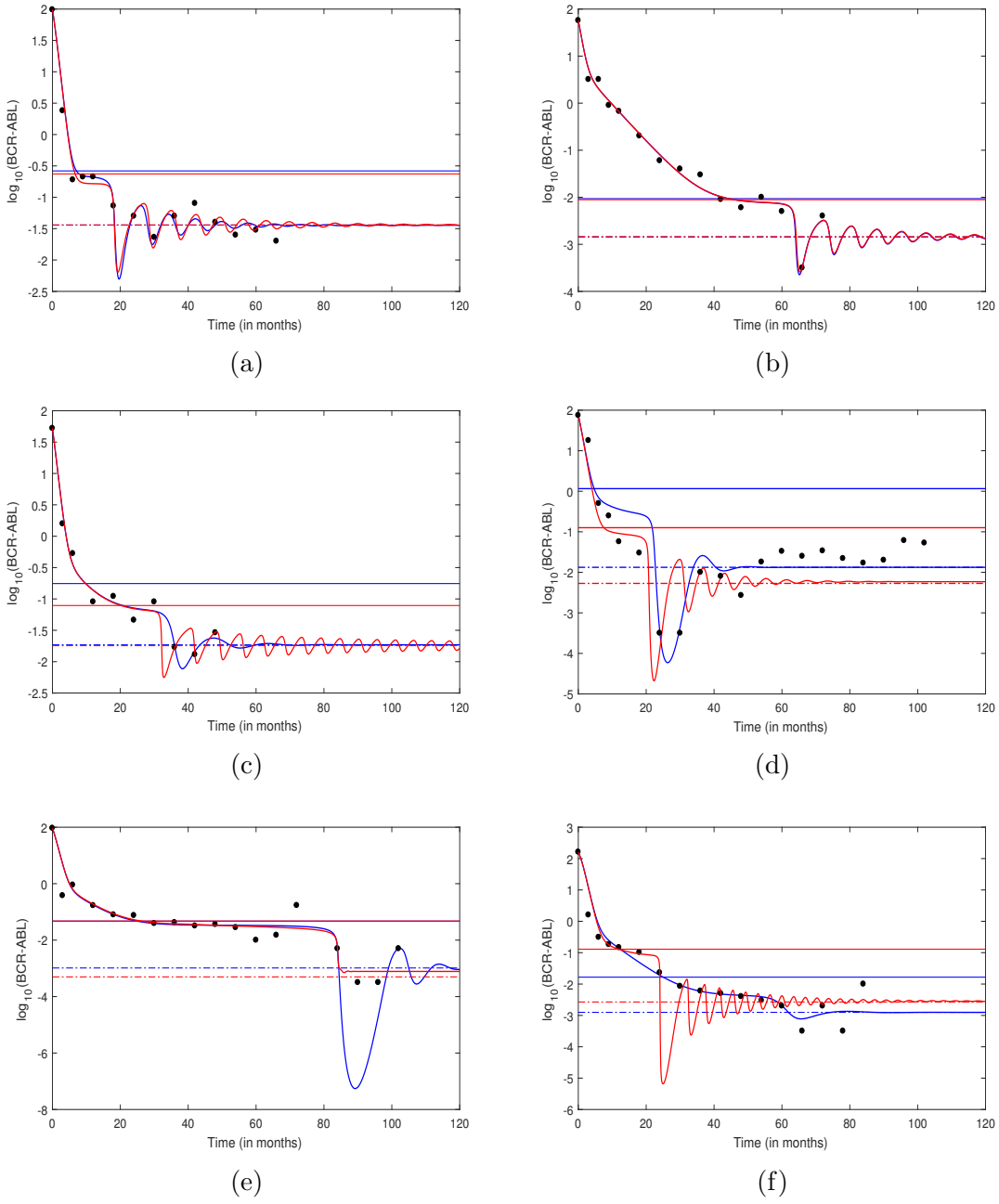


Figure 3.6: Fits of the Clapp model (blue) and our extended model (3.1a)-(3.1g) (red) to patient data for six different patients. The log of the BCR-ABL ratio is plotted over time. Dots represent patient data. Horizontal lines show the immune windows for the respective models.

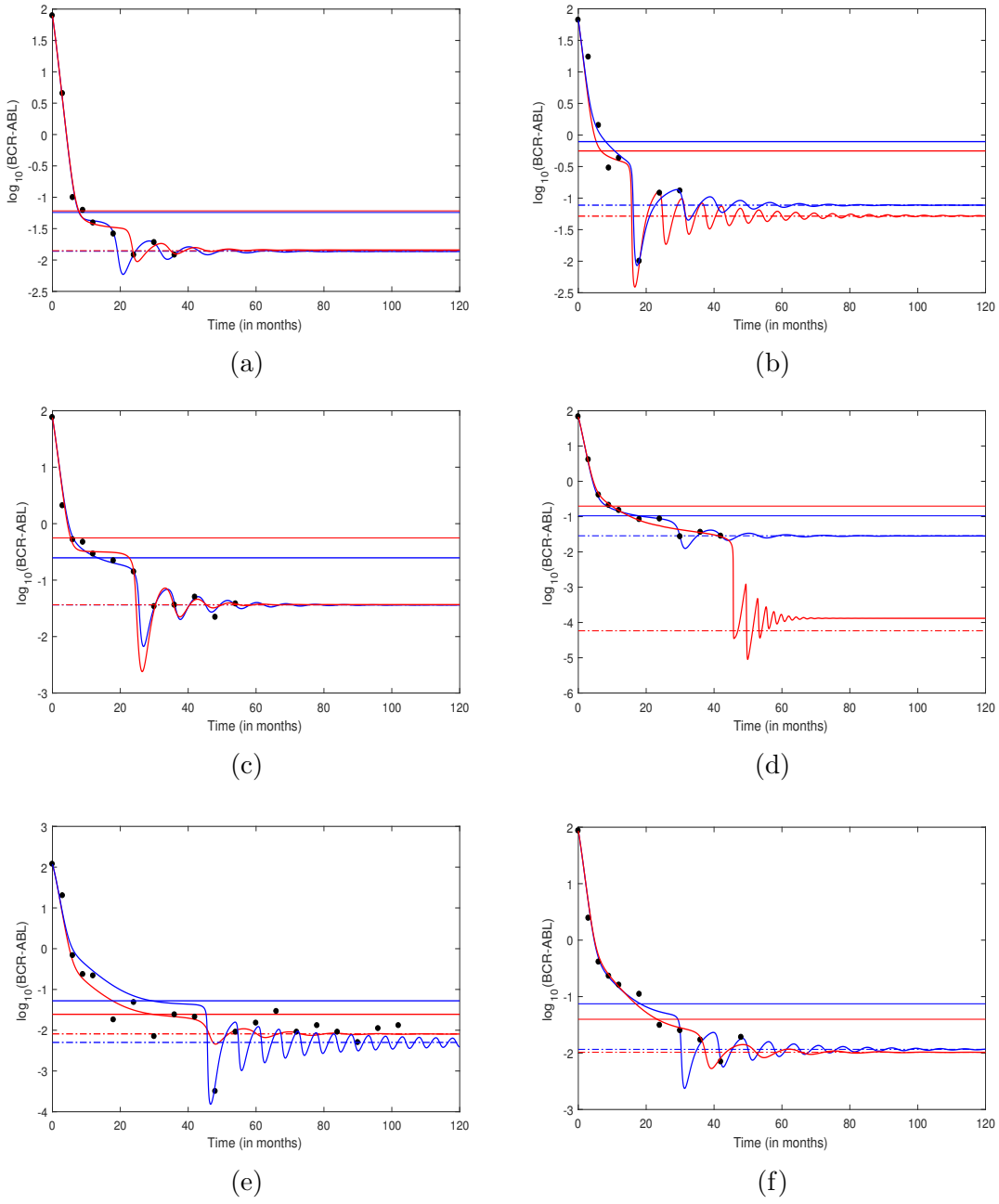


Figure 3.7: Fits of the Clapp model (blue) and our extended model (3.1a)-(3.1g) (red) to patient data for six additional patients. The log of the BCR-ABL ratio is plotted over time. Dots represent patient data. Horizontal lines show the immune windows for the respective models.

response by regulatory cells and lack of activation factors allows for a regrowth of the leukemic population. It is reasonable to then ask if reduction or removal of regulatory cells is enough to counteract this regrowth.

To answer this question, we looked at the effect of removing regulatory cells from the extended model by setting the Treg differentiation rate (r_T) to zero. For roughly 40% of patients in our data set, there is no change in the dynamics of the BCR-ABL ratio (fig. 3.8a). Our model suggests that for the remaining 60% of patients, modulation of regulatory cells may improve outcomes. Changes in dynamics include a smaller minimum value of the BCR-ABL ratio after the initial decline (fig. 3.8d), increased frequency of oscillations in the BCR-ABL ratio (fig. 3.8c) and the possibility of a cure (fig. 3.8b). While it is not clinically feasible to completely eliminate regulatory T cells, these results suggest that an attempt to reduce the number of or the function of these cells may be an important addition to CML treatment protocol.

Since the effects of imanitib are generally the same as all TKI therapies, this extended model of CML and immune response should easily be able to describe the dynamics of CML under second and third generations of the drug and therefore could be fit to data from dasatinib and nilotinib patients. Comparison of patient-specific parameters between these groups of patients may be beneficial in identifying key characteristics between the three iterations of the drug. Of more interest to furthering the study of CML through mathematical models, would be to further expand the description of the immune system in this model. Splitting the effector T cell compartment into CD4 and CD8 T cells would allow for better description of the

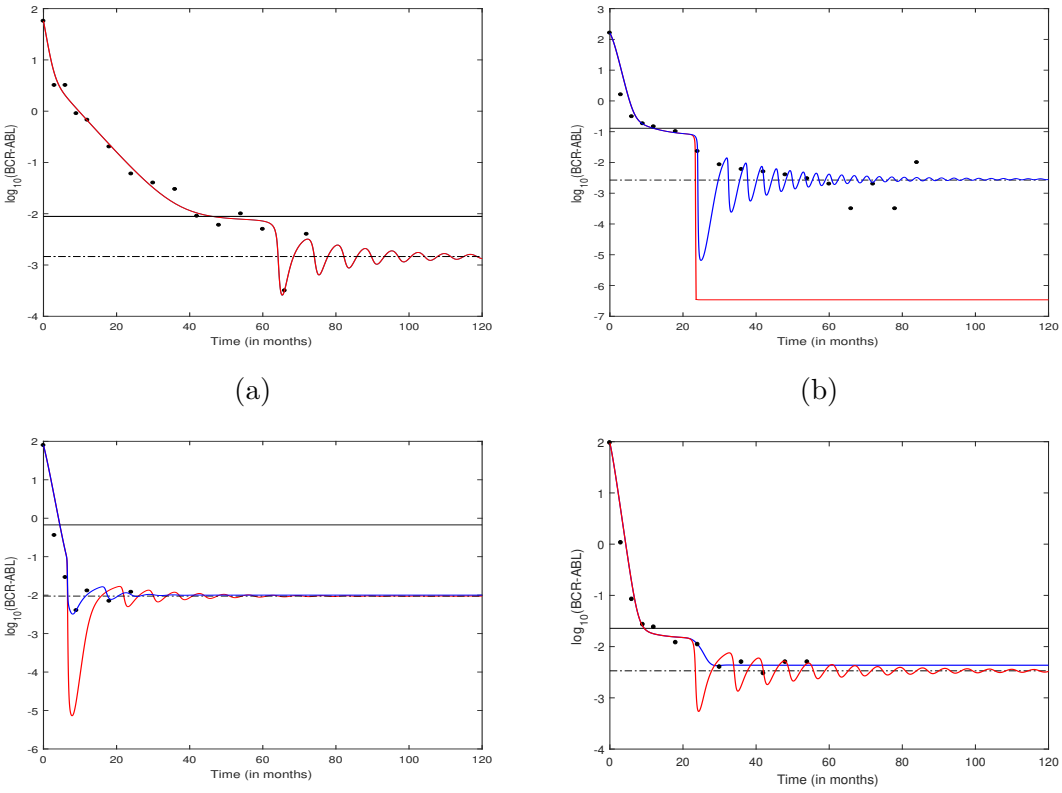


Figure 3.8: Effect of removing regulatory T cells from the model for four representative patients. Simulation of the log BCR-ABL ratio using full extended model shown in blue and without regulatory T cells ($r_T = 0$) shown in red. Removing regulatory cells had various effects: (a) no change in treatment outcome, (b) a simulated cure, (c) more dramatic initial decline in BCR-ABL ratio (d) increased oscillatory behavior.

formation of regulatory T cells. Inclusion of antigen presenting cells, natural killer cells or other innate immune cells that have been shown to be present during CML therapy could allow for the exploration of other immunomodulatory treatments. With these additions, the effects of IFNα, could be incorporated to aid ongoing clinical research into the resurgence of this drug allowing our model to be useful in determining combination therapy protocol.

Chapter 4: Adaptive Parameters

4.1 Introduction

The extended model of CML and the immune response (3.1) provides an approximation of the BCR-ABL ratio over time for a given patient. However, there is no reason to assume that this simulation represents the true dynamics of CML under imatinib therapy. A major reason for this uncertainty is the identifiability of the parameter values, particularly with respect to the patient-specific parameters. There is little evidence to suggest that these values must remain fixed over the course of treatment. In fact, biologically at least some of these values may be likely to change over time. For example, inh_1 and inh_2 may decrease over time if TKI therapy becomes less effective or if treatment is gradually terminated. While immune systems of healthy people are generally stable over time, this is not necessarily true for cancer patients. For instance, patients living with chronic diseases, like CML, can experience T cell exhaustion, a state of the immune system defined by low effector cell function [33]. Such a condition could be reflected in a change in immune parameters later on in the treatment phase. The goal of this chapter is to examine parameter values over shorter windows of data to determine which parameters must change, and in what way, for the model to have the most predictive

power.

This goal is approached in the following way. Working from the theory that the human immune system for a single person has little variation over an extended time span [28], it is hypothesized that changes in parameter values will not be large in magnitude and will most likely change gradually over the treatment time frame. Consider a window of length l being shifted by s months, a subset of the complete treatment time frame. Beginning with the window $[0, l]$ months, the selected parameter values are fit to the data points contained in that window. The window is then shifted to $[s, l + s]$ months and parameters are refit to data points contained in the new window. This process is continued until the current window contains the last data point in the patient data set. The measured ratios are available through our collaboration with the group of Dr. Nicolini (Lyon). For the remainder of this thesis, parameter values that are fit to the complete data set for a given patient are referred to as optimal parameters while parameters fit on a given window are referred to as adaptive parameters.

In Section 4.2, we consider different fitting windows and variations of the cost function. This approach is then used to look at TKI treatment cessation data and explore the differences between CML patients that relapse and those that maintain TFR. A summary of these simulations and results is provided in Section 4.3. Concluding remarks are made in Section 4.4.

4.2 Procedure and Results

Using the values listed in Table 3.1 as initial conditions for the ODE solver and optimal parameters as initial conditions for the selected optimization routine, the extended model (3.1) is fit to data in the first window $[0, l]$. The simulated BCR-ABL ratio is computed from the ODE solution and the selected optimization scheme then searches for parameter values that minimize the 2-norm error between the simulated ratio and the data. Once optimal parameters are found for window $[0, l]$, the window is shifted to $[s, l + s]$ and the process is repeated.

After each shift, the initial conditions for both the ODE solver and the optimization routine are reset using the solution of the model and the adaptive parameters from the previous window. We choose to reset the optimization initial conditions rather than use the optimal parameter values on every window. The idea being that parameter values may change gradually over the full simulation time frame, and therefore the parameter values in the next window will be closer to the current adaptive parameter values.

A couple of optimization routines were considered which have various advantages in relation to our adaptive fitting scheme. Matlab's *lsqnonlin*, a nonlinear least-squares solver, uses a trust region reflective algorithm to minimize the sum of squares of the input function. Advantages of this routine include the ability to bound the parameter values. However, *lsqnonlin* relies on the use of finite differences to estimate the derivative of the objective function, in absence of a provided gradient. Since the function we seek to optimize depends on the numerical solution

of our ODE model, a number of issues can arise when using finite difference approximations including amplification of inaccuracies in the numerical methods and insensitivity of the model to small changes in parameter values. Therefore, matlab's *fminsearch* is considered a better alternative. This routine uses a direct search method to minimize the objective function, but does not accommodate additional constraints on the solution. To avoid optimal solutions that are invalid in the context of our application, for example parameter values that are negative, the value of the objective function is set to NaN.

When determining the length of the fitting window, the amount of data available for fitting and the number of total windows are considered. The length of each window needs to be short enough for the total data set to be covered by multiple windows, but long enough that there are enough data points in each window to provide a reasonable fit. Since the majority of patients in the data set have measurements covering at least 3 years of treatment, a window length of 2 years is selected. As data is available every 3 or 6 months, this guarantees at least 5 data points in each fitting window.

Three different window shifts are considered: 6 months, 1 year, and 2 years. The shorter shifts provide for more overlap in windows and allow us to consider how the model reacts to each newly available data point. A 2 year shift eliminates all overlap in windows and can be thought of as looking at different phases of CML under imatinib therapy. For many patients the initial biphasic decline in BCR-ABL ratio occurs within the first two years of therapy and could therefore be completely contained in the first window, isolated from the evolving dynamics in the later phases

of treatment.

This adaptive fitting approach is applied to the data described in Chapter 3. Results for four example patients are shown in Figures 4.1 and 4.2 for the three different shift values. Refitting patient-specific parameters in the first window of [0, 24] months does not seem to alter the simulation of the model, suggesting that the extended model provides a fairly consistent description of the dynamics of CML in the first phase of treatment. The 6 month and 1 year shifts provide slightly more variation in predictions over these first two years, but generally the dynamics are quite similar. Behavior over the remaining windows seems to vary by patient. For certain patients, the adaptive parameters increase the amplitude and frequency of oscillations in the BCR-ABL ratio (figs. 4.1a and 4.1b), while for others the adaptive parameters smooth out oscillations in the ratio (fig. 4.2b 6 month and 1 year). Occassionally, the adaptive parameters match the dynamics produced by the optimal parameters (fig. 4.2a and fig. 4.2b 2 year). In most cases the adaptive fits produce reasonable alternative simulations of the dynamics of CML, that is, they match the data as well as the optimal fits as measured by the 2-norm errors. There are a few patients where the adaptive fits significantly reduce error. These patients typically have a BCR-ABL ratio that tends to increase, patient 4 (fig. 4.1b) for example.

How the adaptive parameters change from the optimal values and across the windows depends on the patient and the window shift. Figure 4.3 shows the percent of change in each parameter value as compared to the optimal parameter value for each of the four example patients using a 2 year shift. There is no pattern among

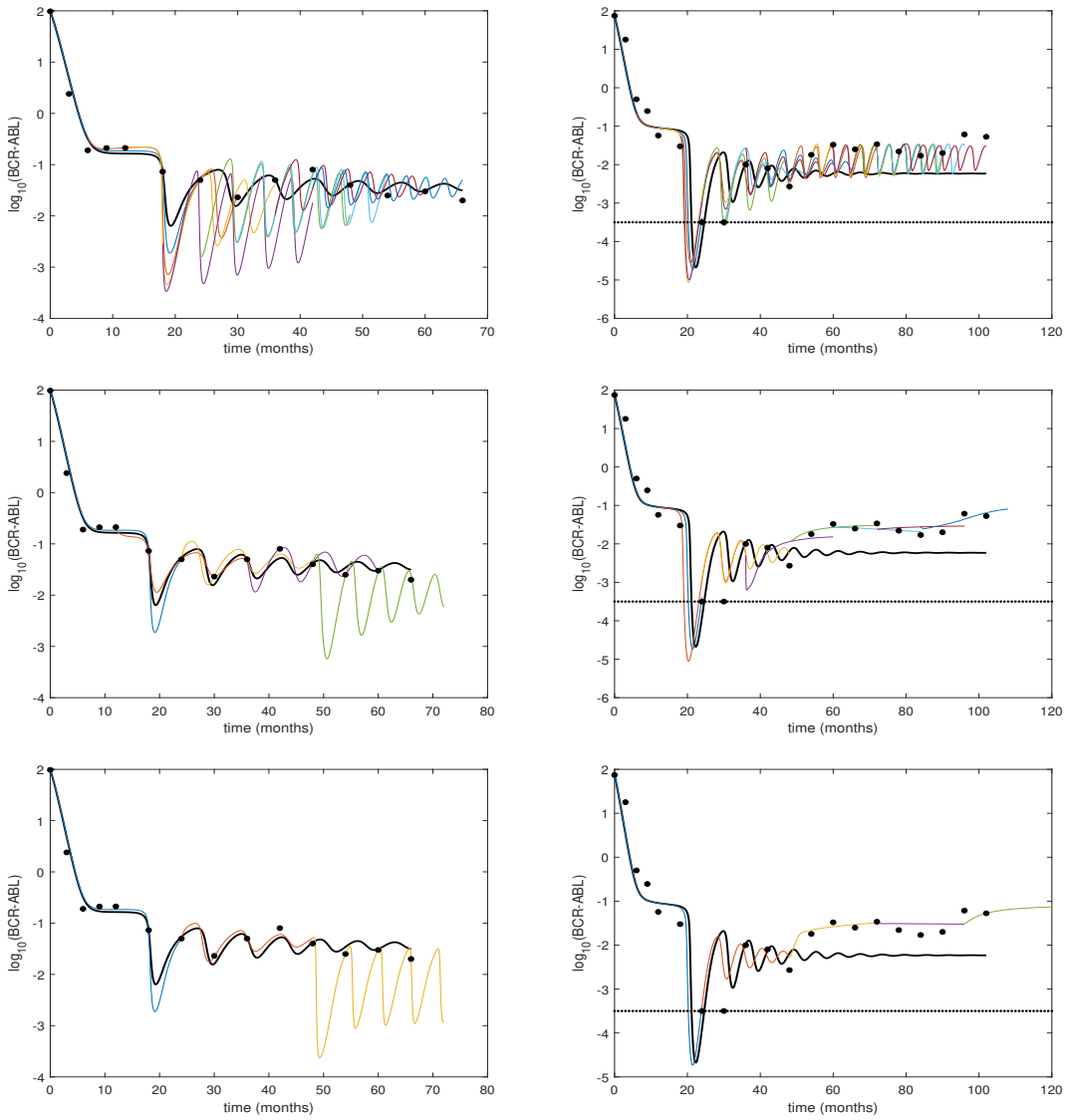
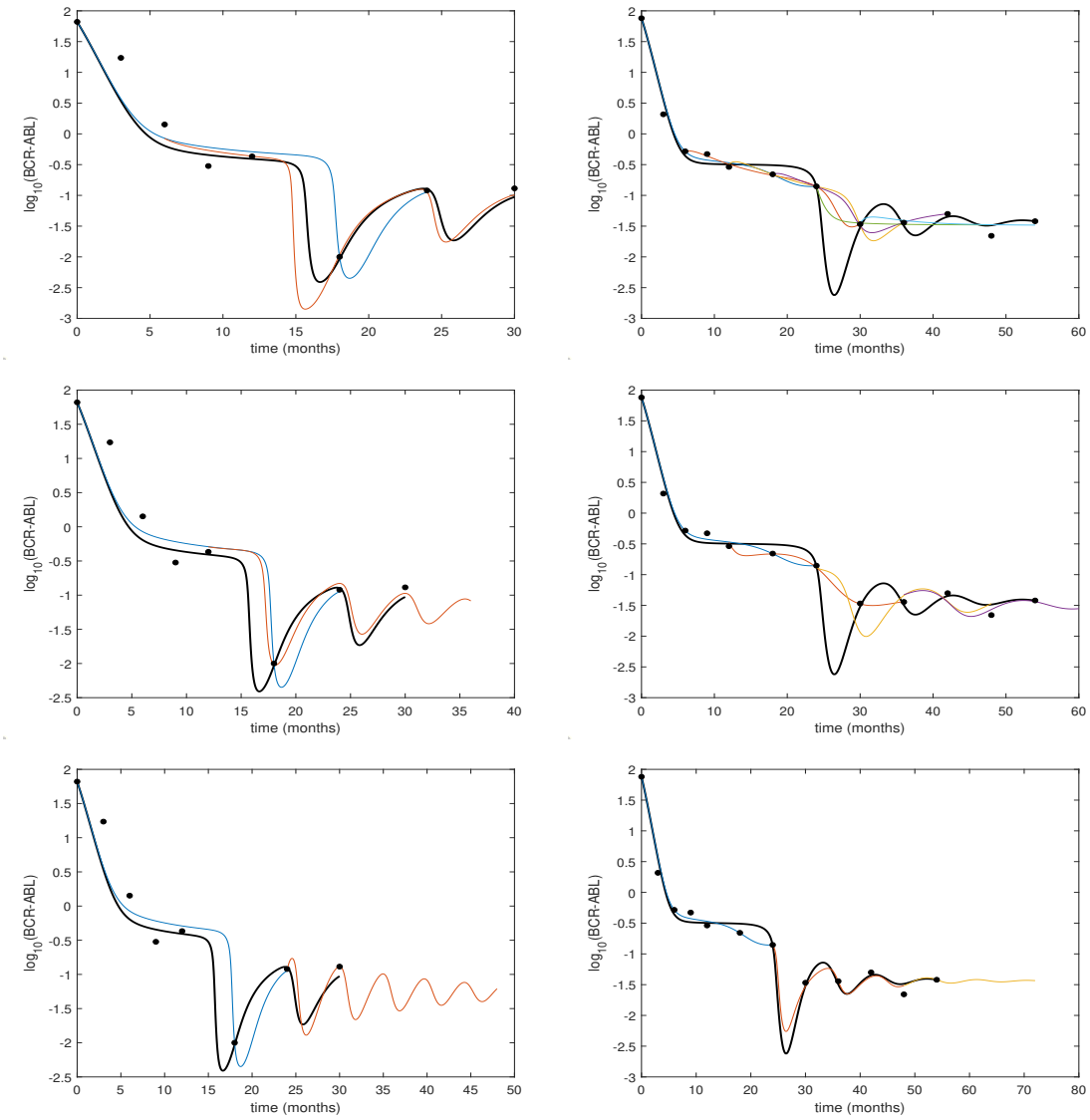


Figure 4.1: Adaptve fits of the the extended model (3.1a)-(3.1g) to patient data for Patient 1 (a) and Patient 4 (b) on windows with 6 month, 1 year and 2 year shifts (top to bottom). The log of the BCR-ABL ratio is plotted over time. Dots represent patient data. The black curve show simulation of the model over the full time span using optimal parameter values (table 3.2). Colored curves indicate simulation over each window. Black horizontal lines indicate the minimal level detectable by RT-PCR. Points are this line correspond to those points at which the disease was not observed.



(a) (b)

Figure 4.2: Adaptive fits of the the extended model (3.1a)-(3.1g) to patient data for Patient 8 (a) and Patient 9 (b) on windows with 6 month, 1 year and 2 year shifts (top to bottom). The log of the BCR-ABL ratio is plotted over time. Dots represent patient data. The black curve show simulation of the model over the full time span using optimal parameter values (table 3.2). Colored curves indicate simulation over each window. Black horizontal lines indicate the minimal level detectable by RT-PCR. Points are this line correspond to those points at which the disease was not observed.

the four patients, with certain parameters increasing for one patient, decreasing for another, and exhibiting very little change in value for others. If we assume that these are accurate reflections of the true characteristics of patient disease and immune response, then these differences between patients further confirm the variability of immune response between patients and suggest that these parameter values are indeed patient-specific. Despite these vast differences, three parameter values governing immune response tend to be highly variable across all patients: maximal effective kill rate of immune cells μ, supression rate of regulatory cells k_T and the differentiation rate of effector T cells in regulatory cells r_T. These values seem to differ greatly from the corresponding optimal parameter values even when the resulting simulations are quite similar, as seen in Patients 1 and 9 when using the 2 year shift, suggesting an identifiability issue. The average percent change in parameter values for all patients is displayed in Figure 4.4. It is clear that the variations in parameter values depend on the shift in the window as these three plots show little similarities.

Immune systems of different patients can vary greatly. However, the immune profile of a single patient is not likely to vary greatly over time. And yet, the adaptive fitting procedure often produces large changes in certain parameter values. Since the selected optimization scheme cannot handle constraints on the solution, we add a second term to the objective function that forces adaptive parameters to remain close to the initial parameters

$$objective = ||log(BCR - ABL_{sim} - log(BCR - ABL_{meas})||_2 + c \cdot ||p_{adapt} - p_0||_2$$

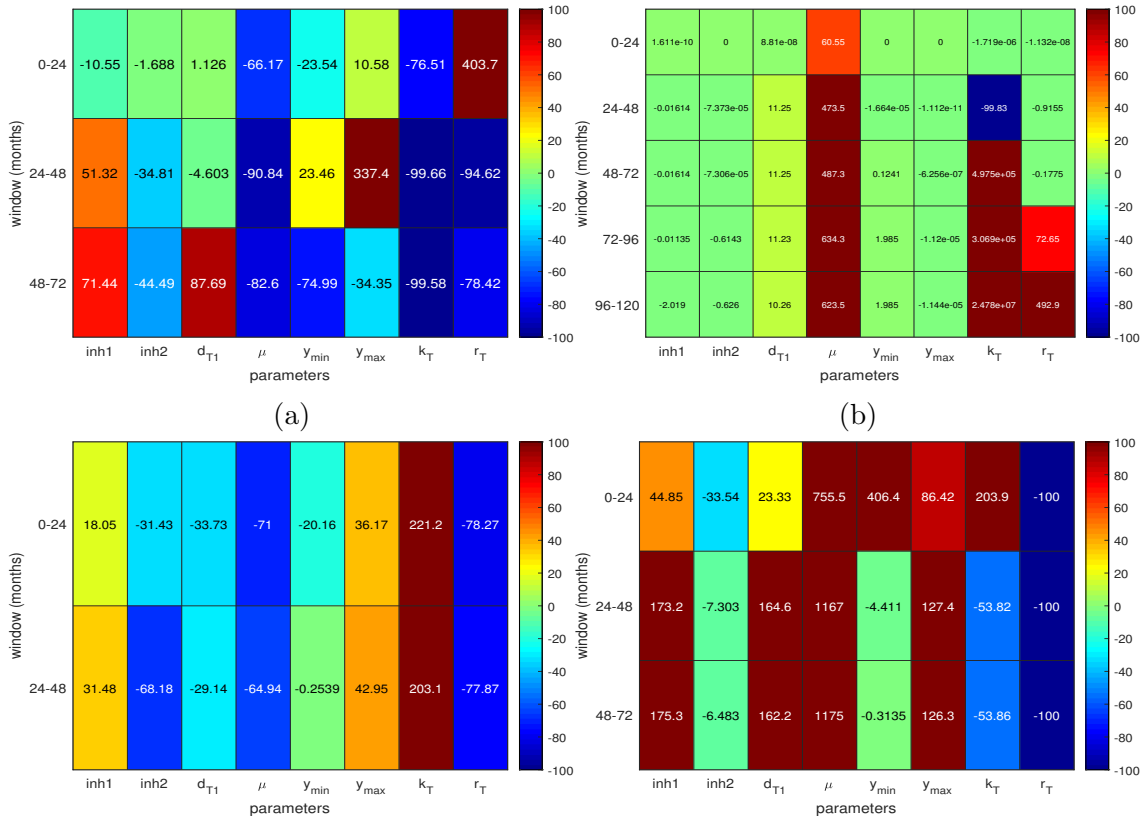
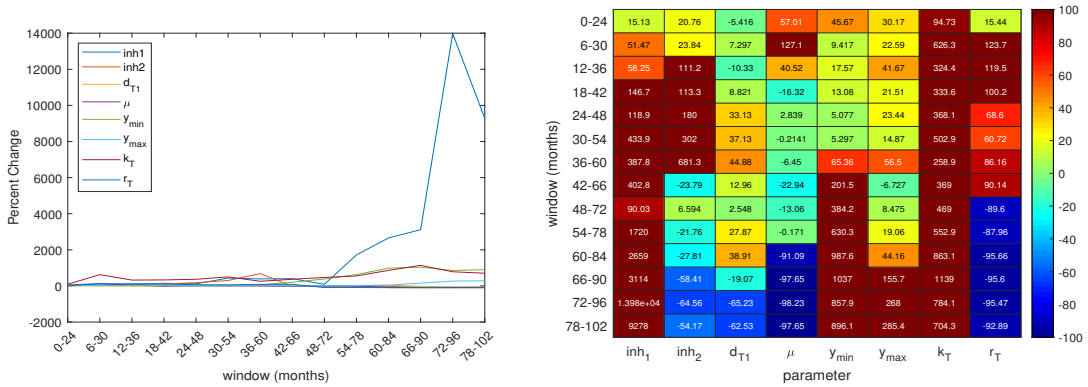
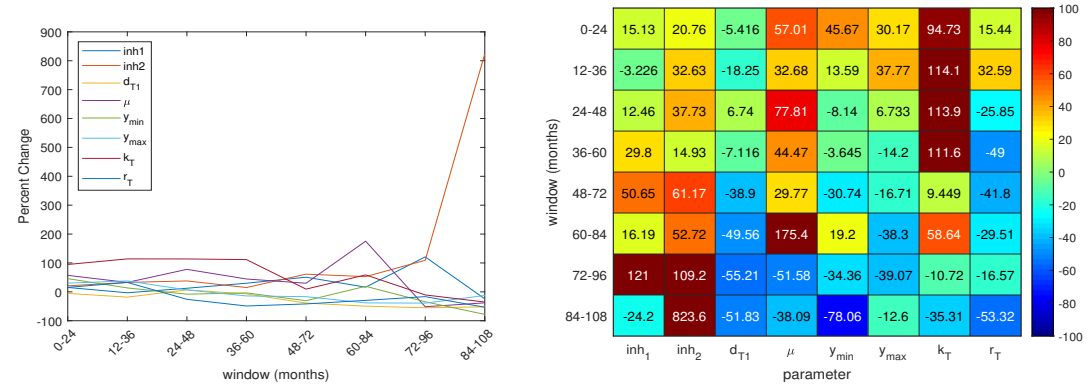


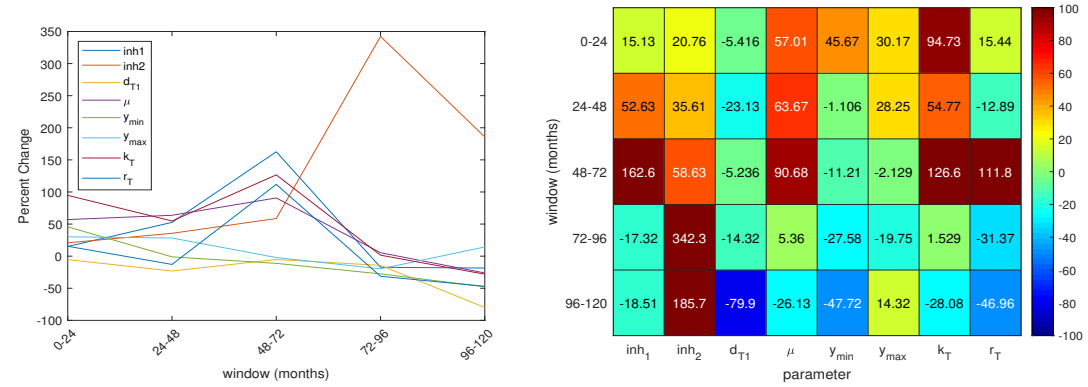
Figure 4.3: Percent change in value of the adaptive parameters from the optimal parameter values for four example patients. Adaptive parameter values are for windows with a 2 year length and 2 year shift. Cells are colored from a 100% decrease (dark blue) to a 100%$^+$ increase (dark red) in parameter value. (a) Patient 1, (b) Patient 4, (c) Patient 8, (d) Patient 9.



(a) 6 month shift



(b) 1 year shift



(c) 2 year shift

Figure 4.4: Average percent change in parameter values for each of the three window shifts (a) 6 months, (b) 1 year, (c) 2 year. Cells of the heat maps are colored from a 100% decrease (dark blue) to a $100\%^+$ increase (dark red) in parameter value.

and consider different weights $c \in [0, 1]$.

Adding this second penalty (section 4.2) reduces the change in parameter values over time, while still improving the quality of fit in most cases. A few simulations for an example patient are provided in Figure 4.5. In general, the dynamics do not differ greatly as the weight on the second penalty or the size of the window shift change. Despite this, certain parameter values do seem to change over the different windows. This behavior is summarized in Figure 4.6. Parameter values governing the effect of imatinib (inh_1, inh_2) and the immune window (y_{min}, y_{max}) rarely change value when the second penalty is added to the cost function. Of these four, inh_1 changes the most over the windows, but never by more than 20%. The remaining four parameters show much more dramatic behavior. The death rate of effector T cells d_{T1} generally steadily increases over the windows, suggesting that T cells die off more quickly as the length of treatment increases. The maximal kill rate of effector cells (μ) generally increases over the first half of windows and then steadily decreases. The values of this parameter are much larger than the optimal value during the first half of treatment and then much smaller than the optimal value in the later phases of treatment. Similarly, the adaptive value of r_T is on average increased from the optimal value on the first half of windows and decreased on the second half, although this parameter fluctuates in value more than μ. This could suggest that both μ and r_T are dependent on the phase of treatment, and that the optimal fit produces some average value for the parameters. Lastly, the regulatory cell suppression rate k_T has behavior that varies depending on the weight c. In general, this parameter fluctuates over treatment. Again, this could suggest that

the value of k_T is dynamic, or simply that it is unidentifiable.

4.3 Application to Treatment Cessation Data

In this section, we explore the ability of our model to capture the dynamics of CML for patients who are candidates for treatment cessation. Further, we seek to use the method of adaptively fitting parameters on smaller windows of data to determine any differences in the underlying dynamics or in the parameter values between patients who experience relapse and those that maintain TFR. We are particularly interested in the windows of data that occur just before treatment cessation. Determining any qualitative or quantitative difference between the two types of patients could be vital in aiding successful treatment cessation.

Data for this section are unpublished observations during and after TKI therapy and are provided by our collaborator, Dr. Nicolini, Centre Léon Bérard. The data set contains BCR-ABL ratios for 79 patients over a follow-up period of at least 33 months. All patients in the study had achieved MMR while undergoing TKI therapy, most had also achieved CMR as noted by at least one undetectable BCR-ABL ratio measurement. BCR-ABL ratios were measured at diagnosis as well as every 3 months for the first year of therapy followed by every 6th month thereafter, including throughout the cessation period. When plotted, the data suggests that patients in this study experienced the same general behavior in BCR-ABL ratio. Within the first two years of therapy, a noticeable biphasic decline in the ratio occurs, followed in most cases by damped oscillations approaching a low disease steady

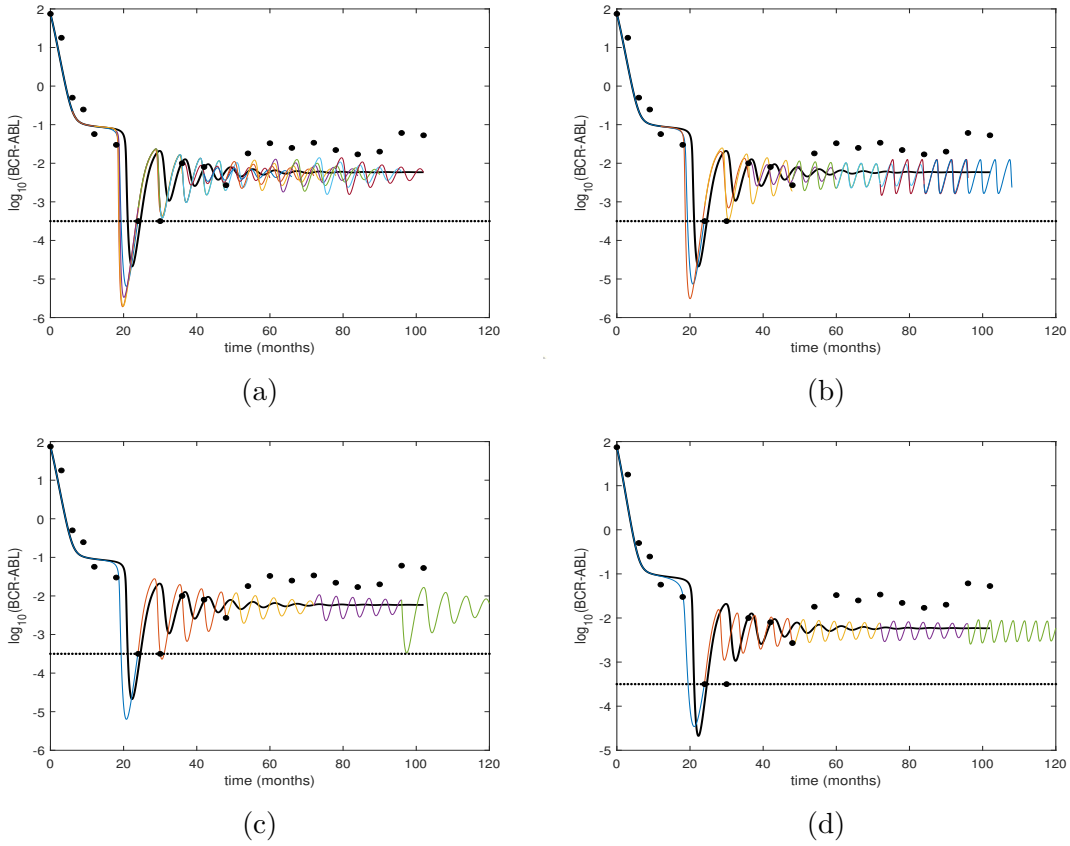


Figure 4.5: Adaptive fit of the the extended model (3.1a)-(3.1g) to patient data for Patient 4 with second penalty (section 4.2) added to the cost function. The log of the BCR-ABL ratio is plotted over time. Dots represent patient data. The black curve shows simulation of the model over the full time span using optimal parameter values (table 3.2). Colored curves indicate simulation over each window. Black horizontal lines indicate the minimal level detectable by RT-PCR. Points are this line correspond to those points at which the disease was not observed. (a) 6 month window shift and $c = 1$, (b) 1 year window shift and $c = 0.75$, (c) 2 year window shift and $c = 0.5$, (d) 2 year window shift and $c = 0.25$

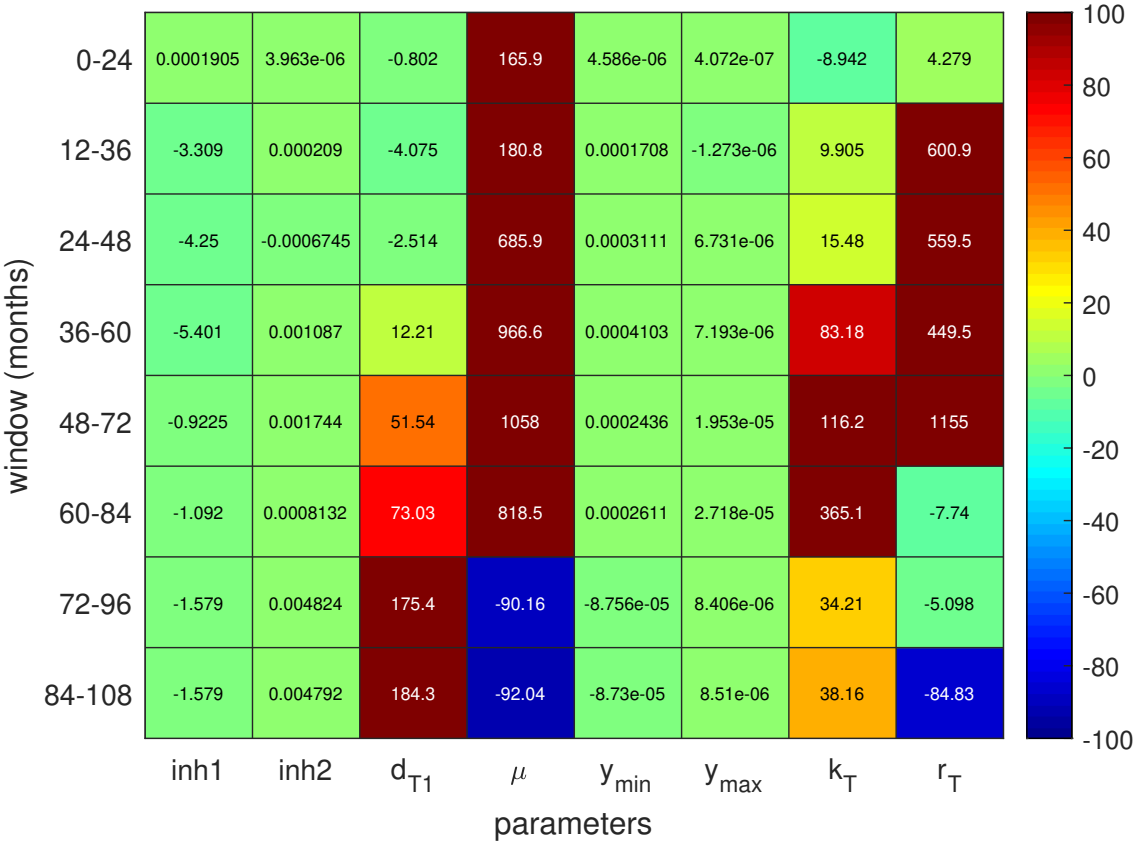


Figure 4.6: Average percent change in parameter values for a window shift of 1 year and second penalty (section 4.2) weighted by $c = 0.75$. The average percent change is similar for all window shifts and second penalty weights. Cells of the heat maps are colored from a 100% decrease (dark blue) to a $100\%^+$ increase (dark red) in parameter value.

state. The noticeable difference in this data set is the high number of undetectable measurements. On average, each patient has 8 undetectable measurements across the course of treatment and following treatment cessation. Despite the large number of undetectable measurements, it is expected that the extended model will still capture the underlying dynamics of the disease.

Of the 79 patients in the database, 30 had taken IFN prior to or in combination with a TKI. Imatinib was the primary TKI of use (47 patients). The remaining patients used second generation TKIs, dasatinib (11) or nilotinib (21). Since our model does not take the effects of IFN into account, we considered only patients who were currently using imatinib and had never taken IFN. Approximately 2/3 of these remaining 29 patients did not experience a relapse during the study. A relapse is indicated by loss of MMR and the restarting of TKI therapy. After resumption of therapy, all patients regained MMR.

The extended model of CML and immune response (eq. (3.1)) was fit to data for each of these patients, using the process described in Section 3.3.1 to obtain patient-specific parameter values. Because the underlying dynamics of treatment cessation are not fully captured by this model, only data points occurring before the primary treatment cessation date were used for fitting. Initial conditions and universal parameter values are as listed in Table 3.1. The logarithm of the simulated BCR-ABL ratio for a select 15 patients is plotted over the course of the study, up to the first treatment cessation date, as shown in Figures 4.7 and 4.8. The patient-specific parameter values are provided in Table 4.1. The selected example patients are an even mix of relapse and TFR cases, and each have at least 5 detectable BCR-

ABL measurements prior to cessation to improve the ability of fitting the model. From the simulations (figs. 4.7 and 4.8), it is evident that the model captures the general dynamics of CML prior to TKI cessation. Thus the model performs well, reasonably depicting the dynamics of CML during imatinib therapy.

There does not appear to be any discernible difference between the predicted dynamics of relapse and remission patients. A summary of parameter values for both relapse and TFR patients is provided in Table 4.2. It appears that on average inh_1 and inh_2 are higher for relapse patients, while y_{max} and k_T are much larger in value for TFR patients. This would suggest that imatinib plays a greater role in the reduction of cancer load for relapse patients. A larger immune window for TFR patients, indicated by a larger y_{max} value, would mean that an immune response begins sooner in the treatment phase for these patients. It could also indicate that relapses post-treatment cessation would take a longer period of time to occur, if at all. The values of k_T and r_T also appear to be larger on average for TFR patients while μ is smaller. This is interesting as this would suggest that T cells are less effective at killing leukemic cells and that the average immune response in TFR patients is effected more by regulatory cells. It is important to note that these averages are calculated from small samples and may be biased by a single extreme value in one patient. Figure 4.9 better represents the distribution of patient-specific parameter values in both relapse and TFR patients.

We choose not to refit parameter values at treatment cessation due to the scarcity of the data. Patients that relapse tend to lose MMR quickly and restart TKI therapy soon after, resulting in only one to two data points in this time frame.

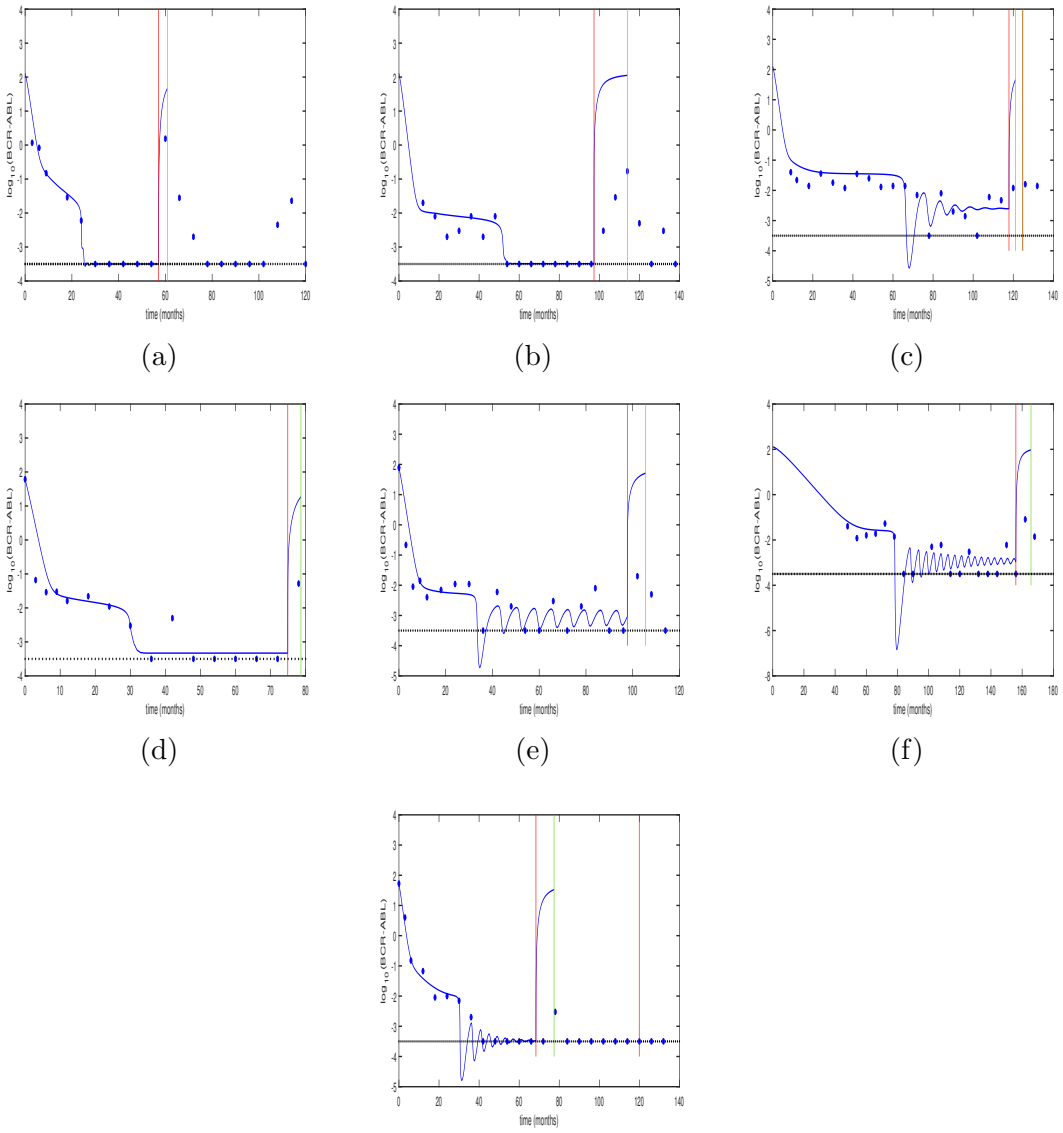


Figure 4.7: Fits of the the extended model (3.1) to patient data for seven patients who experienced relapse after treatment cessation. The log of the BCR-ABL ratio is plotted over time. Dots represent patient data. Black horizontal lines indicate the minimal level detectable by RT-PCR. Points are this line correspond to those points at which the disease was not observed. Red vertical lines indicated time of treatment cessation, while green vertical lines indicate a restart of treatment.

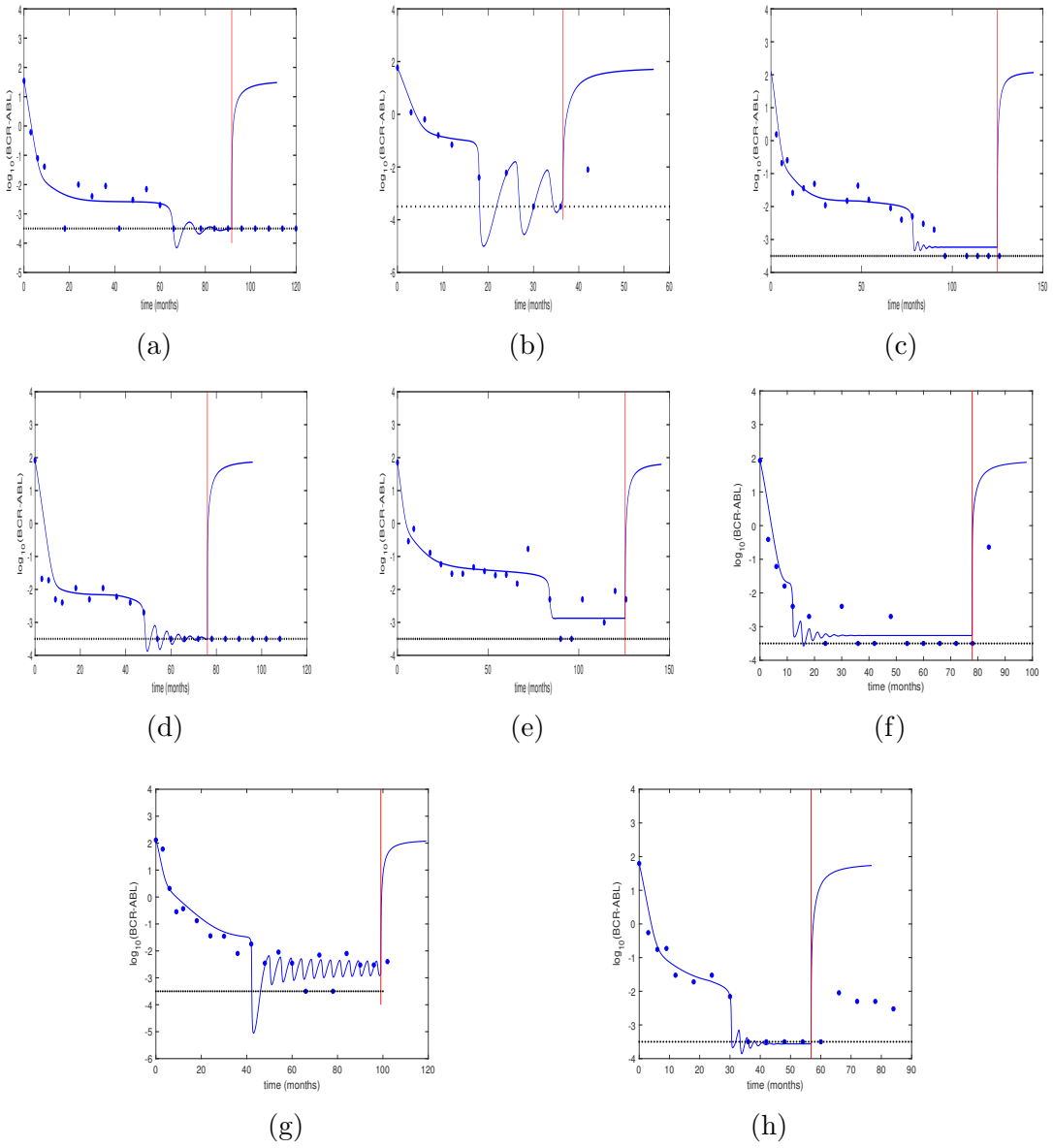


Figure 4.8: Fits of the the extended model (3.1) to patient data for eight patients with TFR after treatment cessation, at least until end of study. The log of the BCR-ABL ratio is plotted over time. Dots represent patient data. Black horizontal lines indicate the minimal level detectable by RT-PCR. Points are this line correspond to those points at which the disease was not observed. Red vertical lines indicated time of treatment cessation.

Patient	inh_1	inh_2	d_{T1}	μ	y_{min}	y_{max}	k_T	r_T
1	39.187	413.995	0.100	$2.120 \cdot 10^{-10}$	93.564	$7.263 \cdot 10^{4}$	1.050	$7.473 \cdot 10^{-11}$
2	2.3763	$1.012 \cdot 10^{4}$	0.072	$4.622 \cdot 10^{-10}$	90.8323	$1.204 \cdot 10^{4}$	$1.539 \cdot 10^{-4}$	$6.769 \cdot 10^{-7}$
3	8.210	769.773	0.031	$1.042 \cdot 10^{-7}$	$1.664 \cdot 10^{3}$	$2.738 \cdot 10^{4}$	8.990	$1.577 \cdot 10^{-12}$
4	2.556	$1.731 \cdot 10^{3}$	0.058	$2.220e \cdot 10^{-15}$	99.7233	$3.384 \cdot 10^{5}$	0.114	$1.636 \cdot 10^{-14}$
5	4.227	$4.775 \cdot 10^{3}$	0.122	$6.918 \cdot 10^{-7}$	$1.307 \cdot 10^{3}$	$8.238 \cdot 10^{3}$	0.095	$2.245 \cdot 10^{-7}$
6	$8.371 \cdot 10^{3}$	1.043	0.091	$7.230 \cdot 10^{-7}$	784.265	$1.960 \cdot 10^{4}$	0.257	$2.722 \cdot 10^{-9}$
7	26.510	278.965	0.064	$1.074 \cdot 10^{-8}$	661.7145	$4.360 \cdot 10^{4}$	$4.497 \cdot 10^{-6}$	$6.001 \cdot 10^{-5}$
8	17.481	864.572	0.025	$1.812 \cdot 10^{-8}$	968.778	$1.332 \cdot 10^{4}$	12.625	$8.038 \cdot 10^{-12}$
9	4.912	186.556	0.011	$2.212 \cdot 10^{-9}$	$1.165 \cdot 10^{3}$	$1.228 \cdot 10^{6}$	2.429	$7.214 \cdot 10^{-12}$
10	31.215	488.623	0.065	$6.489 \cdot 10^{-10}$	238.149	$1.300 \cdot 10^{4}$	$1.586 \cdot 10^{-4}$	$3.584 \cdot 10^{-7}$
11	4.111	$4.071 \cdot 10^{3}$	0.023	$1.522 \cdot 10^{-9}$	220.378	$1.624 \cdot 10^{4}$	$4.859 \cdot 10^{-8}$	$5.768 \cdot 10^{-4}$
12	24.546	81.071	0.095	$1.756 \cdot 10^{-10}$	793.277	$1.617 \cdot 10^{5}$	9.7834	$8.274 \cdot 10^{-12}$
13	1.430	$4.660 \cdot 10^{3}$	0.049	$7.900 \cdot 10^{-10}$	357.739	$1.080 \cdot 10^{5}$	48.102	$1.160 \cdot 10^{-12}$
14	140.724	54.703	0.180	$1.142 \cdot 10^{-7}$	$1.534 \cdot 10^{3}$	$2.734 \cdot 10^{4}$	0.001	$6.556 \cdot 10^{-7}$
15	20.013	253.240	0.066	$4.700 \cdot 10^{-10}$	290.783	$8.824 \cdot 10^{4}$	1.123	$3.864 \cdot 10^{-11}$

Table 4.1: Patient parameter values used for the simulations in Figure 4.7 and Figure 4.8. Patients 1-7 correspond to Figures 4.7a to 4.7g. Patients 8-15 correspond to Figures 4.8a to 4.8h. The bounds of the immune window are fit to the data rather than the parameters α and ϵ. These parameters can be computed from the immune window as $\epsilon = 1/(y_{min}y_{max})$ and $\alpha = (y_{min} + y_{max})\epsilon d_{T1}$.

		inh_1	inh_2	d_{T_1}	μ	y_{min}	y_{max}	k_T	r_T
Relapse	Min	2.376	1.043	0.031	$2.220 \cdot 10^{-15}$	90.832	$8.238 \cdot 10^{3}$	$4.497 \cdot 10^{-6}$	$2.245 \cdot 10^{-7}$
	Mean	1207.674	2583.881	0.077	$2.186 \cdot 10^{-7}$	671.560	$7.456 \cdot 10^{4}$	1.501	$8.638 \cdot 10^{-6}$
	Max	8370.650	10117.917	0.121	$7.230 \cdot 10^{-7}$	1663.632	$3.384 \cdot 10^{5}$	8.989	$6.001 \cdot 10^{-5}$
	STD	3158.607	3703.000	0.030	$3.361 \cdot 10^{-7}$	632.370	$1.184 \cdot 10^{5}$	3.323	$2.265 \cdot 10^{-5}$
Remission	Min	1.430	54.703	0.011	$1.756 \cdot 10^{-10}$	220.377	$1.299 \cdot 10^{4}$	$4.858 \cdot 10^{-8}$	$1.160 \cdot 10^{-12}$
	Mean	30.554	1332.411	0.064	$1.727 \cdot 10^{-8}$	696.015	$2.070 \cdot 10^{5}$	9.258	$7.223 \cdot 10^{-5}$
	Max	140.724	4659.758	0.180	$1.142 \cdot 10^{-7}$	1533.830	$1.228 \cdot 10^{6}$	48.102	$5.768 \cdot 10^{-4}$
	STD	45.765	1896.488	0.054	$3.963 \cdot 10^{-8}$	495.758	$4.162 \cdot 10^{5}$	16.433	$2.039 \cdot 10^{-4}$

Table 4.2: The minimum, mean, maximum and standard deviation are computed from the patient-specific parameter values of the relapse and remission patients in the Stop TKI data set.

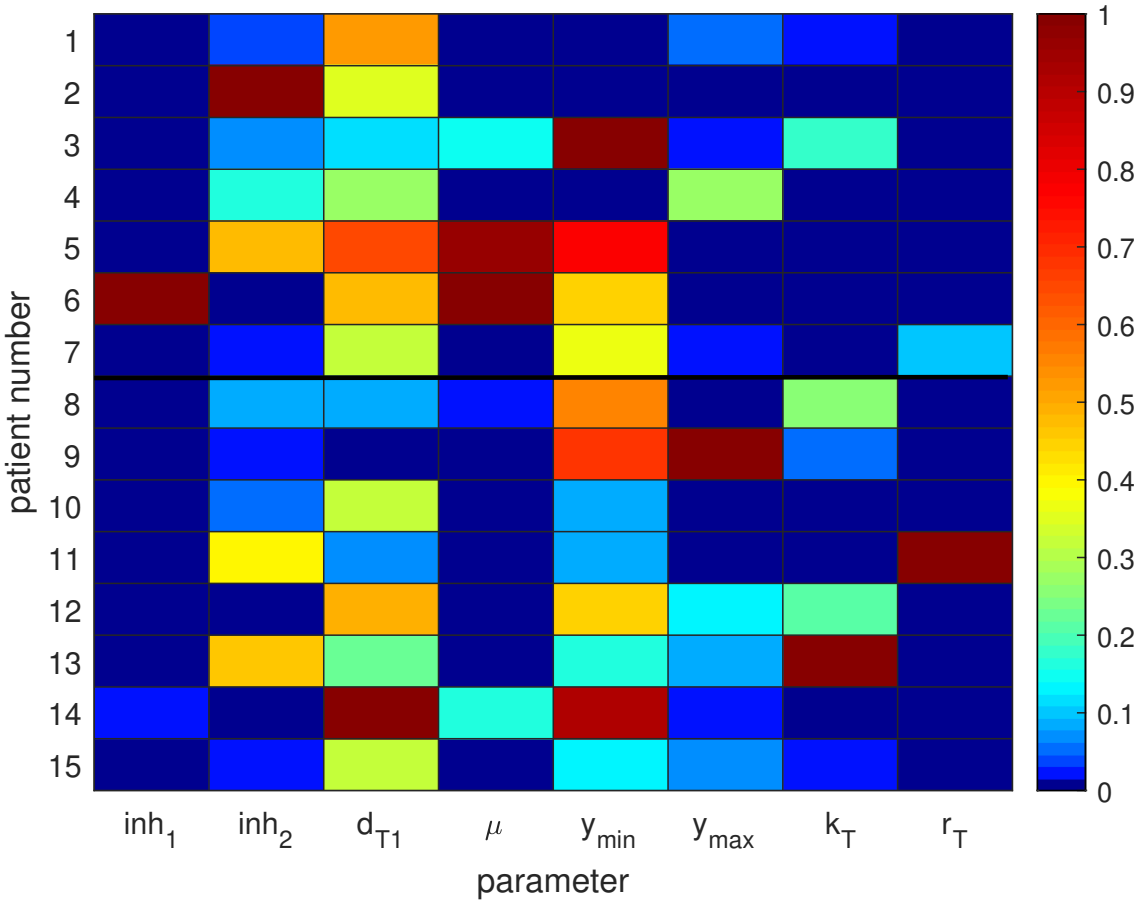


Figure 4.9: Heat map of patient-specific parameter values. Colors in each cell correspond to the parameter values in Table 4.1 normalized by column. Larger values displayed in dark red. The smallest values are in dark blue. Relapse and remission patient-specific parameters are separated by a horizontal line.

Instead, following treatment cessation, the parameter values inh_1 and inh_2 are set to a value of 1, indicating that leukemic cells regain their natural amplification and differentiation rates with loss of TKI inhibition. The system is then simulated until therapy is restarted or for a period of 20 months if resumption does not occur. The extended model ((3.1)) predicts a relapse for all 15 patients, including the 8 TFR patients that never restart imatinib therapy within the span of this study. This suggests an inability of the model to provide an accurate prediction of CML progression after therapy under the assumptions that leukemic cells fully regain their proliferative ability and that the immune system experiences no changes.

With no clearly discernible difference in the dynamics or the patient-specific parameter values between relapse and remission patients, we next look to the adaptive fitting process to see if any distinctions arise between the two classes of patients. All 15 Stop TKI patient data was adaptively fit using 6 month, 1 year and 2 year windows and the second penalty ((4.2)) weighted by various values. The dynamics for each patient do not generally change as the length of the shift and the weight on the penalty change. Example adaptive fits are found in Figure 4.10 for four patients, two relapse and two remission. As with the previous data set, the adaptive fit agrees with the optimal fit on the first window [0, 24], with small deviations seen only in a couple patients. The adaptive method is more suited to improving the quality of fit in the later stages of treatment, where the two resulting simulations often differ greatly. However, looking at the simulations alone, there again does not appear to be any noticeable difference in the dynamics between relapse and TFR patients.

How the parameter values adapt to each window provides slightly more insight.

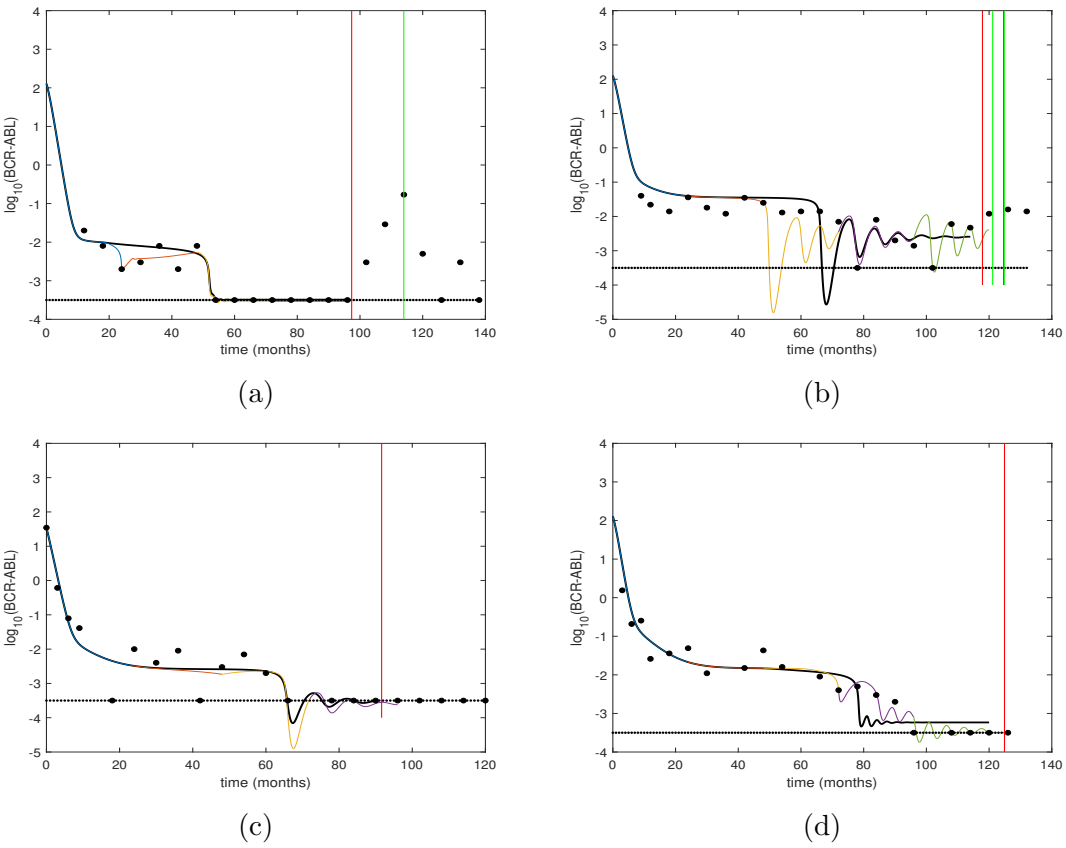


Figure 4.10: Adaptve fits of the the extended model (3.1a)-(3.1g) to Stop TKI patient data for (a) Patient 2, (b) Patient 3, (c) Patient 8, and (d) Patient 10 on windows with a 2 year shift and second penalty weighted by $c = 0.75$. The log of the BCR-ABL ratio is plotted over time. Dots represent patient data. The black curve show simulation of the model over the full time span using optimal parameter values (table 4.1). Colored curves indicate simulation over each window. Black horizontal lines indicate the minimal level detectable by RT-PCR. Points are this line correspond to those points at which the disease was not observed. Treatment cessation and resumption are indicated by vertical red and green lines, respectively.

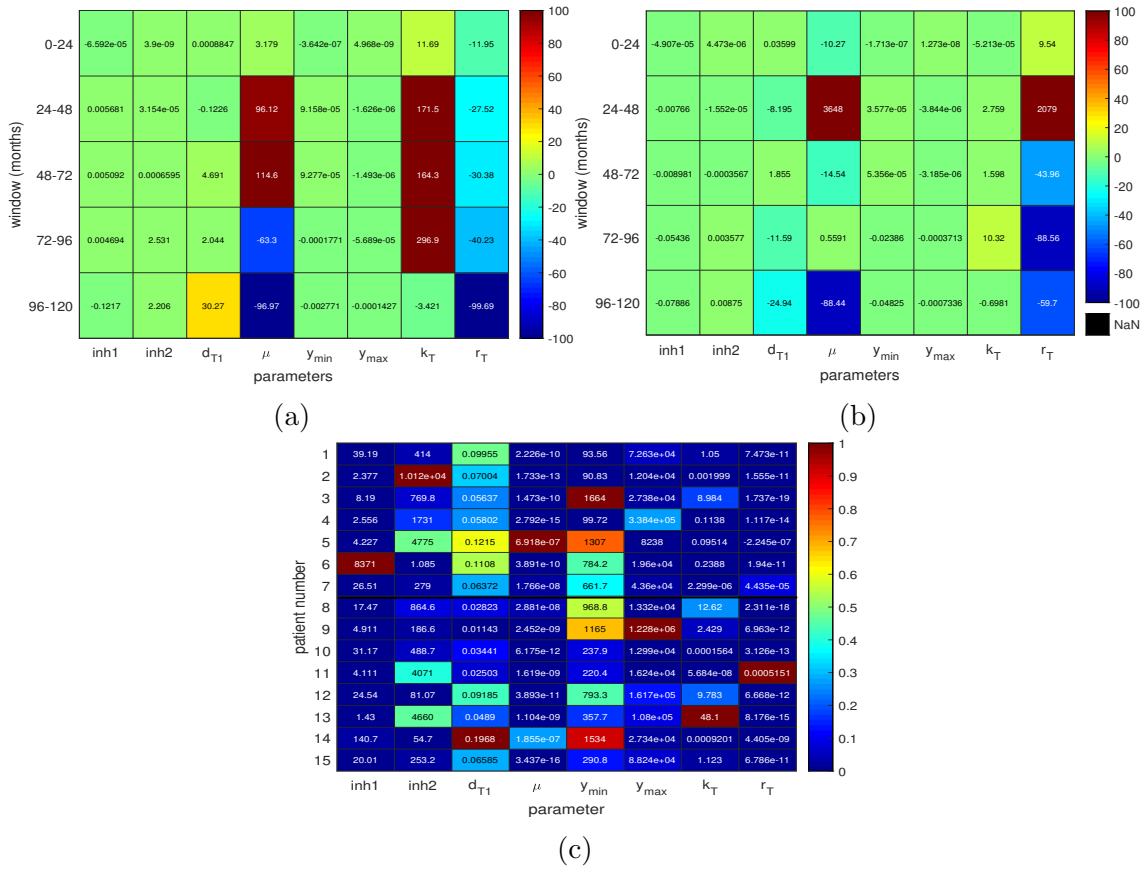


Figure 4.11: Average percent change in parameter values for (a) relapse patients and (b) remission patients using a window shift of 2 years and second penalty weighted by $c = 0.75$. (c) Heat map of patient-specific parameter values on the 2 year window leading up to treatment cessation. Cells of (a) and (b) are colored from a 100% decrease (dark blue) to a $100\%^+$ increase (dark red) in parameter value. Cells of (c) are colored by parameter value and normalized by column. Larger values displayed in dark red. The smallest values are in dark blue. Relapse and remission patient-specific parameters are separated by a horizontal line.

For both relapse and TFR categories, the average percent change in parameter value from the optimal was computed for each combination of window shift and second penalty weight. Figures 4.11a and 4.11b give a comparison of these changes for a 2 year window and a weighted second penalty of $c = 0.75$. Parameter values for inh_1, inh_2, y_{min} and y_{max} are fairly consistent across the different windows for both relapse and TFR patients. For this specific window shift and weight combination, the behavior of the other four patient-specific parameters differs between relapse and TFR.

Parameter values for relapse patients tend to follow certain trends. For example, d_{T1} gradually increases in value over the windows, reaching its largest value prior to cessation. The maximal kill rate μ increases over the first half of treatment and then decreases in the later phase. There is also a large increase in the value of k_T, which suggests an increase in self-regulation. These trends suggest that the immune response is weakened closer to treatment cessation and could provide explanation for relapse. In TFR patients, μ and r_T tend to fluctuate in value over the windows, while d_{T1} gradually decreases and k_T experiences very little change.

It should be noted that Figure 4.11a is representative of the average percent change for all window shifts and weights c for relapse patients, i.e. the shift and penalty do not greatly affect the adaptive fit. However, this is not the case for TFR patients. The drug and immune window parameters are consistently fixed in the adaptive fit, but there are no discernible trends in the other four immune parameters for TFR patients. This could mean that no trend exists for these parameters, or perhaps that some of these patients are misclassified as TFR.

Lastly, we examined the parameter values themselves in the window immediately preceding treatment cessation. Although many of the parameters have changed from the optimal values for each patient, the distribution of values is quite similar to that of the optimal (fig. 4.9 vs. fig. 4.11c). Relapse and TFR patients cannot be categorized by their parameter values alone.

4.4 Discussion

The dynamics of CML under imatinib therapy can be split into two phases. The first phase is a well-documented biphasic exponential decline in the BCR-ABL ratio. This characteristic has been widely noted in many CML patients and there are numerous mathematical models that capture this behavior. Chapter 3 introduced a new model to capture not only this initial decline, but also the oscillatory nature of the BCR-ABL ratio in later phases of treatment. Due to these two phases and the possibility of a varying immune response, we hypothesized that the patient-specific parameters in our model, governing the effect of imatinib and the immune response, were likely to vary in value and adapt to the phase of therapy. Therefore, an adaptive data fitting process was developed in which the model parameters are refit on shifting windows of data. A window length of 2 years was selected and various window shift lengths were explored. In order to prevent parameter values from extreme changes between windows, a second penalty term was added to the objective function in the optimization scheme. We looked at not only how the simulated dynamics differed between the adaptive and optimal fits, but also in what

way the parameter values changed.

The goal set out at the beginning of this chapter, to improve the predictability of the model through adaptively fitting to the data, was not met. The adaptive fitting method is not perfect. Issues arise in using an optimization routine in combination with a "function" that is an estimated solution of an ODE. Variations in our results could be attributed to the inability of the model to react to small changes in parameter values. We also recognized the possibility of overfitting our model. This is not a big data problem but rather a scarce data problem, as there are very few data points in each window. Overfitting is not a large concern if the the purpose of the model is to qualitatively understand the dynamics of CML and the immune system during imatinib therapy. However, it is a concern that will need to be accounted for if the model is to be used for forecasting CML progression. Lastly, we are assuming the data is completely accurate, ignoring measurement and representation errors in our data fitting process. These errors should be taken into account if the model is to be used to quantitatively represent CML.

Despite these imperfections in the adaptive fitting process, we do gain insights into CML and immune response and are able to determine what may be needed to improve the forecasting ability of the extended model. First, a few trends became apparent for certain parameter values that give some insight into disease dynamics. The large variability in immune parameters (d_{T1}, μ, k_T, r_T) throughout treatment suggest that identifying or controlling these values may be key to fully understanding the role of immune response, and therefore successful treatment, for CML patients. While no clear distinction was made between relapse and TFR patients when using

the adaptive fitting method on Stop TKI data, it was observed that immune parameters of relapse patients signaled a decrease in immune response in the stages of therapy immediately preceding treatment cessation. Further exploration of immune characteristics in these patients, including immune assays directly measuring the immune response, could be vital to fully understanding the cause of relapse in CML patients. Second, the adaptive fit tends to provide differing yet equally plausible dynamics from the optimal simulation, especially in the oscillatory phase of therapy, suggesting that not enough is known about the underlying dynamics in this phase. Further revisions or extensions of the model may be necessary. Additional data types would be beneficial in accomplishing this. Finally, new methods of parameter estimation are necessary to improving the confidence in parameter values and thus model predictions.

Chapter 5: Data Assimilation Methods

5.1 Introduction

Data assimilation is a technique that combines a dynamical model of a physical process with observations of that process to determine the optimal state of the system. Numerical models can provide general descriptions of the underlying dynamics of a physical process, but are imperfect due to numerical errors, uncertainty in initial conditions and unknown parameter values. Thus, they often lack the ability to accurately quantify the present or future state of the system. Observations of the state are also imperfect. Available data are often sparse and are accompanied by measurement errors. Additional error in observations can arise due to inaccurate representation of the observations as a function of the state variables. Data assimilation seeks to account for all of these errors, combining the dynamical and statistical properties of the system, to improve predictability of the model.

The longest documented use of data assimilation techniques are in meteorology where the goal is to provide reliable weather forecasts by merging well-adapted atmospheric models with noisy observations of the state (temperature, velocity, etc.). Data assimilation has also become standard practice in oceanography, improving understanding of ocean dynamics. Other applications of data assimilation to the

geosciences, including seismology and modeling of Earth's magnetic field, naturally gained popularity. More recently, data assimilation has been gaining traction in biomedical applications, where observations come in numerous forms including electrocardiograms, MRIs, CT scans, and cell-based assays. Specific applications of data assimilation techniques in medical applications include brain tumor growth [34], cardiac function [35, 36], epidemiology [37] and neurology [38]. Nonetheless, this is a relatively novel approach to studying disease progression and response to treatment, specifically in relation to CML. Successful use of data assimilation in these fields could launch major developments in personalized medicine. The use of data assimilation in mathematical oncology could be particularly beneficial [39].

We have shown that the extended model of CML and the immune system (3.1) has advantages over previous models of CML under TKI therapy, capturing not only the typical biphasic decline of the BCR-ABL ratio but also oscillations in the ratio that are experienced by numerous patients. In addition, the extended model better represents the immune response in CML patients by incorporating self-regulation and a more biologically accurate representation. However, we have also demonstrated that the model is only a general description of the underlying dynamics of the disease. Simplifications of biological processes, estimated initial conditions, and unknown patient-specific parameter values lead to uncertainties in the prediction of the model. Applying data assimilation methods to models of CML would take into account errors in model formulation and model noise to improve forecasts, and ideally, allow these mathematical models to provide valid predictions in a clinical setting.

Most data assimilation techniques are based on a Bayesian framework. A given dynamical model provides a projection of the state variable X, often unobservable. It is assumed that the prior distribution $p(x)$ is known or can be reasonably estimated. Noisy observations Y can be thought of as conditioned on the true state x, and so $p(y|x)$ describes the distribution of measurement errors [40]. Bayes' Theorem states that the posterior distribution $p(x|y)$ is proportional to the product of the prior and measurement distributions. Knowledge of the posterior distribution can then be used to update X. Often data assimilation techniques such as the Kalman filter and its extensions, as well as Monte Carlo methods like sequential Monte Carlo and particle filters, are based in Bayesian inference. The choice of a particular technique can depend on whether the available dynamical model is linear or nonlinear and what is known about the distribution of errors.

5.2 The Extended Kalman Filter

The Extended Kalman Filter (EKF) is selected for this application, since the extended model of CML and the immune system (3.1) is nonlinear. The Kalman filter provides the optimal estimate for linear systems. This nonlinear extension, which is commonly used for nonlinear models with Gaussian error distributions, relies on a local linearization of the model.

As with the Kalman filter, the EKF consists of two main steps: prediction and update. Let $\boldsymbol{x}$ be an N-dimensional vector describing the state of a system or physical process. Assume that the true system is governed by a stochastic nonlinear

differential equation

$$\dot{\boldsymbol{x}^t} = \boldsymbol{f}(\boldsymbol{x}^t) + \dot{\boldsymbol{\eta}^t}, \tag{5.1a}$$

$$\boldsymbol{\eta}^t \sim \mathcal{N}(0, \boldsymbol{Q}^t). \tag{5.1b}$$

Here $\boldsymbol{\eta}^t$ is the process noise, normally distributed with mean zero and $N \times N$ covariance matrix $\boldsymbol{Q}^t$. The superscript 't' indicates the true state. Suppose observations $\boldsymbol{y}^o$ become available at times t_i for $i = 1, ..., p$. It is assumed that observations are imperfect, accompanied by Gaussian white noise with mean zero and covariance matrix $\boldsymbol{R}_j^t$.

$$\boldsymbol{y}_i^o = \boldsymbol{y}_i^t + \boldsymbol{\epsilon}_i, \tag{5.2a}$$

$$\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \boldsymbol{R}_i^t). \tag{5.2b}$$

Suppose there are M different types of observations of the system made at a given time t_i. Then $\boldsymbol{y}^o$ is an M-dimensional vector and $\boldsymbol{R}$ is $M \times M$. Observations are related to the state variable through an observation function $\boldsymbol{h}$, possibly nonlinear, such that

$$\boldsymbol{y}_j^t = \boldsymbol{h}_j(\boldsymbol{x}^t(t_j)).$$

The EKF begins with the prediction or forecast step during which both the state vector $\boldsymbol{x}$ and the error covariance matrix $\boldsymbol{P}$,

$$\boldsymbol{P} \equiv \mathbb{E}[(\boldsymbol{x} - \boldsymbol{x}^t)(\boldsymbol{x} - \boldsymbol{x}^t)^T],$$

of the system are forecasted forward in time according to

$$\dot{\boldsymbol{x}^{\boldsymbol{f}}} = \boldsymbol{f}(\boldsymbol{x}^{\boldsymbol{f}}), \tag{5.3a}$$

$$\dot{\boldsymbol{P}^{\boldsymbol{f}}} = \boldsymbol{F}\boldsymbol{P}^{\boldsymbol{f}} + \boldsymbol{P}^{\boldsymbol{f}}\boldsymbol{F}^{\boldsymbol{T}} + \boldsymbol{Q}, \tag{5.3b}$$

where

$$\boldsymbol{F} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\Big|_{\boldsymbol{x}=x^f},$$

is the Jacobian matrix of the deterministic process $\boldsymbol{f}$ and $\boldsymbol{Q}$ is the estimated process noise covariance. The superscript 'f' indicates forecast. The use of $\boldsymbol{F}$, also known as the tangent linear model (TLM), in eq. (5.3) provides a linearization of the nonlinear process $\boldsymbol{f}$ for use with the Kalman filter equations.

At each observation time t_i, the update or analysis step is performed. The goal of this step is to minimize the mean square error J_i,

$$J_i \equiv tr\big(\boldsymbol{P}^{\boldsymbol{a}}(t_i)\big) = tr\Big(\mathbb{E}\Big[\big(\boldsymbol{x}^{\boldsymbol{a}}(t_i) - \boldsymbol{x}^{\boldsymbol{t}}(t_i)\big)\big(\boldsymbol{x}^{\boldsymbol{a}}(t_i) - \boldsymbol{x}^{\boldsymbol{t}}(t_i)\big)^T\Big]\Big).$$

Although no truly optimal solution is computable in finite time for nonlinear dynamics, the EKF consistently provides a first-order approximation [41, 42]. The forecasted variable $\boldsymbol{x}^{\boldsymbol{f}}$ and $\boldsymbol{P}^{\boldsymbol{f}}$ are updated to $\boldsymbol{x}^{\boldsymbol{a}}$ and $\boldsymbol{P}^{\boldsymbol{a}}$ respectively, as follows.

$$\boldsymbol{x}^{\boldsymbol{a}} = \boldsymbol{x}^{\boldsymbol{f}} + \boldsymbol{K}[\boldsymbol{y}^{\boldsymbol{o}} - \boldsymbol{h}(\boldsymbol{x}^{\boldsymbol{f}})], \tag{5.4a}$$

$$\boldsymbol{P}^{\boldsymbol{a}} = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{P}^{\boldsymbol{f}}, \tag{5.4b}$$

where

$$\boldsymbol{K} = \boldsymbol{P^f H^T}(\boldsymbol{H P^f H^T} + \boldsymbol{R})^{-1},$$

is the result of minimizing J, known as the Kalman gain and

$$\boldsymbol{H} = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\Big|_{\boldsymbol{x}=x^f},$$

is the linearization of observation function $\boldsymbol{h}$. The superscript 'a' indicates analysis. The results of the update step, $\boldsymbol{x^a}$ and $\boldsymbol{P^a}$, become initial conditions for the next forecast and the process continues.

5.3 Applying EKF to CML data

To apply the EKF to our model, we consider the state variable $\boldsymbol{x} \in \mathbb{R}^7$ to be the vector of cell concentrations

$$\boldsymbol{x} = [y_0,\, y_1,\, y_2,\, y_3,\, T_0,\, T_1,\, R]^T.$$

The state variable evolves in time according to the forecast model $\boldsymbol{f}$ given by the right hand side of (3.1). To initialize the model error covariance matrix, we first consider

$$\boldsymbol{P}_0 = diag\big((\boldsymbol{x_0} - \boldsymbol{x_0^t})(\boldsymbol{x_0} - \boldsymbol{x_0^t})^T\big).$$

The initial conditions in Chapter 3 are considered suitable based on background information as well as previous models of CML and the immune system. However,

we cannot guarantee that these values are the exact true state $\boldsymbol{x_0^t}$ for each patient. Instead, we estimate reasonable upper and lower bounds for the true state using a 10% margin of error. These upper and lower bounds can then be used to determine an estimate for our initial state $\boldsymbol{x_0} = 0.5(\boldsymbol{x_{ub}} + \boldsymbol{x_{lb}})$ and error covariance matrix where $\boldsymbol{x_0} - \boldsymbol{x_0^t} = 0.5(\boldsymbol{x_{ub}} - \boldsymbol{x_{lb}})$. Additionally, we consider correlation between a few of our state variables based on determination of initial conditions in Chapter 3. Initial concentrations of y_0 and y_1 were computed as percentages of a estimated concentration of total leukemic stem cells. Additionally, the initial concentration of y_2 is determined by considering progenitor cells to be in a steady state ratio with their precursors (y_1) [30]. With this in mind, $\boldsymbol{P_0}$ is structured as a block diagonal matrix where

$$\boldsymbol{P_0} = (\boldsymbol{x_0} - \boldsymbol{x_0^t})(\boldsymbol{x_0} - \boldsymbol{x_0^t})^T, \tag{5.5a}$$

$$\boldsymbol{P}_{i,j} = 0 \quad \text{for} \quad (i,j) \neq (1,2),(2,1),(2,3),(3,2). \tag{5.5b}$$

The non-zero elements of $\boldsymbol{P}_0$ can be further adjusted by multiplying by a factor $k_{i,j}$ to increase or decrease confidence in a particular value.

The process noise covariance matrix $\boldsymbol{Q}$ is difficult to estimate and can have a large effect on performance of the EKF. Many different approaches exist for estimating or tuning $\boldsymbol{Q}$ including Bayesian, maximum-likelihood, covariance matching and correlation techniques. For this work, an adaptive adjustment of $\boldsymbol{Q}$ is applied [43],

updating $\boldsymbol{Q}$ at each analysis step using the innovation

$$\boldsymbol{d} = \boldsymbol{y^o} - \boldsymbol{h}(\boldsymbol{x^f}).$$

Since the true process noise is unknown, it is estimated as

$$\tilde{\boldsymbol{\eta}} = \boldsymbol{x_k^a} - \boldsymbol{f}(\boldsymbol{x_{k-1}^a}) \tag{5.6a}$$

$$= \boldsymbol{K}\boldsymbol{d}. \tag{5.6b}$$

Thus

$$\boldsymbol{Q_{k-1}} = \mathbb{E}\big[\tilde{\boldsymbol{\eta}}\tilde{\boldsymbol{\eta}}^T\big] \tag{5.7a}$$

$$= \mathbb{E}\big[\boldsymbol{K}(\boldsymbol{d}\boldsymbol{d}^T)\boldsymbol{K}^T\big]. \tag{5.7b}$$

The expected value is approximated by taking an average of $\boldsymbol{d}\boldsymbol{d}^T$ over time using a 'forgetting factor' $0 < \gamma \leq 1$, so that at each update step

$$\boldsymbol{Q_k} = \gamma\boldsymbol{Q_{k-1}} + (1-\gamma)\boldsymbol{K}(\boldsymbol{d}\boldsymbol{d}^T)\boldsymbol{K}^T.$$

The value of γ can be adjusted to put more or less weight on previous estimates of $\boldsymbol{Q}$.

For CML patients, a few types of data can be collected. As noted before, BCR-ABL ratios for a set of 104 CML patients have been made available by our collaborator Dr. Franck Nicolini. In the future, it may be possible to addition-

ally obtain cell concentration measurements for various cell types through immune assays. Cell concentrations for both naive and mature CD4+ and CD8+ T cells, as well as regulatory T cells have been measured in numerous studies researching the connection between CML and the immune system. We therefore consider two variations of the observation function $\boldsymbol{h}$. In the case of BCR-ABL ratio alone, $h \in \mathbb{R}$ and is given by (3.2). With the addition of cell concentration data, $\boldsymbol{h} \in \mathbb{R}^4$ and

$$\boldsymbol{h}(\boldsymbol{x}) = [x_5,\, x_6,\, x_7,\, 100\beta\frac{x_4}{2z + x_4}].$$

The last element of $\boldsymbol{h}$ is again (3.2), with x_4 denoting the component of the state variable representing the mature leukemic cell concentration and z denoting an estimate of the total healthy cell population.

Observations are assumed to be uncorrelated, thus $\boldsymbol{R}$ is structured as an $M \times M$ diagonal matrix. In the case of BCR-ABL ratio as the only data type, $\boldsymbol{R}$ is a scalar value. BCR-ABL data is obtained through quantitative PCR. Coefficient of variation (CV) is the typical measurement of inter-assay precision. Low BCR-ABL ratios (<1%) were found to have a CV of 48%, while high BCR-ABL ratios (>1%) have a CV of 17%. These values are quite similar to the CVs between patient duplicates of 50% (BCR-ABL $<$ 1%) and 21% (BCR-ABL $>$ 1%) [8]. Immune assay data for CML patients can be collected by various methods such as enzyme-linked immunosorbent assays (ELISA), fluorescence-activated cell sorting (FACS), and mass cytometry (CyTOF) [44–46]. Assessing accuracy of cellular population measurements produced by flow cytometry is consider impossible as there is no

true standard reference material [47]. Each technique can be affected by various errors, which vary from laboratory to laboratory. When immune assay data is made available by our collaborators, information about measurement errors should also be obtainable. Until then, cellular counts by FACS will be considered to have an imprecision of $\pm$ 5% [48, 49].

Before applying the EKF, the TLM was validated and the assumption of Gaussian error was evaluated to determine if the EKF was a suitable DA technique to use. Since the EKF relies on the linearization of the dynamic model to forecast the model error covariance matrix, it is necessary to determine the accuracy of the TLM. The Taylor expansion of a nonlinear model $\mathcal{F}$ gives

$$\mathcal{F}(\boldsymbol{x} + a\boldsymbol{p}) = \mathcal{F}(\boldsymbol{x}) + a\boldsymbol{p}^T \boldsymbol{F}(\boldsymbol{x}) + O(a^2),$$

where $\boldsymbol{F}$ is the TLM for $\mathcal{F}$, $\boldsymbol{x}$ is the state variable, and $\boldsymbol{p}$ is a perturbation vector. Thus the difference in trajectories of the nonlinear model ($\Delta\boldsymbol{x}$) should be similar to the evolution of the perturbation $\delta\boldsymbol{x}$ by the TLM.

$$\Delta\boldsymbol{x} = \boldsymbol{x}(t_n; \boldsymbol{x_0} + a\boldsymbol{p}, t_0) - \boldsymbol{x}(t_n; \boldsymbol{x_0}, t_0), \tag{5.8a}$$

$$\delta\boldsymbol{x} = \int \dot{\delta\boldsymbol{x}} = \int a\boldsymbol{p}^T \boldsymbol{F}\, dt. \tag{5.8b}$$

These two quantities are computed for a random perturbation $\boldsymbol{p}$ and various scaling

factors a. The relative difference

$$r = \frac{||\Delta \boldsymbol{x} - \delta \boldsymbol{x}||}{||\delta \boldsymbol{x}||},$$

is calculated at each day in a 10 day period. The steps taken here follow the validation approach used by Yang *et al.* [50]. Since the time step used in our model of CML is on the scale of one day, evaluating the relative difference at these points should reasonably inform the validity of our TLM in the EKF formulation. The results are shown in Figure 5.1. The relative difference is quite small for all values of the scaling factor with a minimum value achieved for $a = 10^{-5}$ for all time spans except 1 day. This suggests that the TLM provides a good local linearization of our dynamic model.

To determine the validity of the assumption of Gaussian error in our model, we evolve an ensemble $\boldsymbol{X}$ of 10000 initial states using the extended model and check the distribution at various time points. The ensemble is a multivariate normal with $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{x_0}, \boldsymbol{P}_0)$. The extended model of CML (3.1) was used to evolve the ensemble over a one-year period. The same patient-specific parameter values were used for each member of the ensemble. At each day in the simulation, the Doornik-Hansen omnibus multivariate normality test [51] was used to determine if the forecasted ensemble remained Gaussian. P-values from this test are plotted versus simulation time in Figure 5.2, which suggests that the ensemble is no longer multivariate normal after 11 days at a 0.05 significance level. However, all of the individual cell populations retain a unimodal, roughly normal distribution throughout the simulation

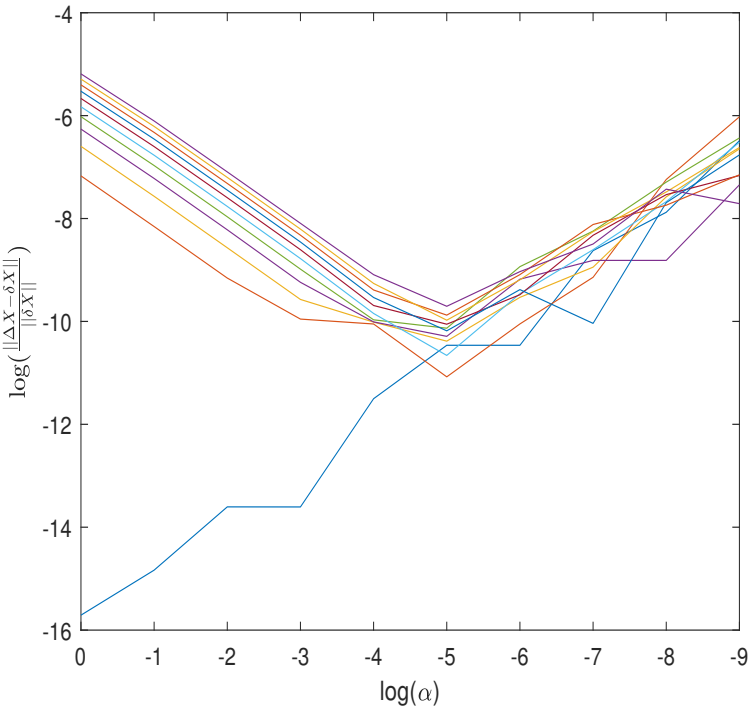


Figure 5.1: The relative difference between $\Delta \boldsymbol{x}$ and $\delta \boldsymbol{x}$ versus the scaling factor a for integration time spans. The one-day integration window increases linearly as the scaling factor decreases. All other integration windows have a minimum relative difference at $a = 10^{-5}$

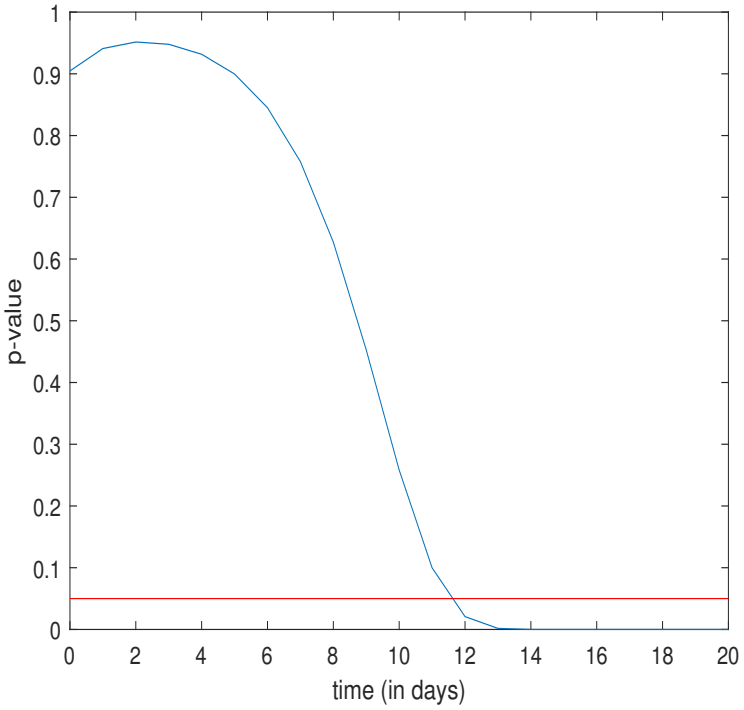


Figure 5.2: P-values for Doornik-Hansen omnibus multivariate normality test for an ensemble of 10000 initial conditions. The ensemble is tested for normality each day in a one-year simulation. Only the p-values for the first 20 days are plotted, after which the p-value is zero. The red line shows the significance level of 0.05.

(fig. 5.3). Since the time between observations is typically at least 3 months, the loss of Gaussianity in a few days could affect the success of the EKF, but together these results are not enough to completely rule out the viability of the EKF.

5.3.1 Numerical Results without Process Noise

In this section, we investigate the performance of the EKF without process noise, i.e. $\boldsymbol{Q} = 0$. BCR-ABL ratio data is assimilated to the extended model of CML and the immune system (3.1) using the EKF for various patients. Parameter values

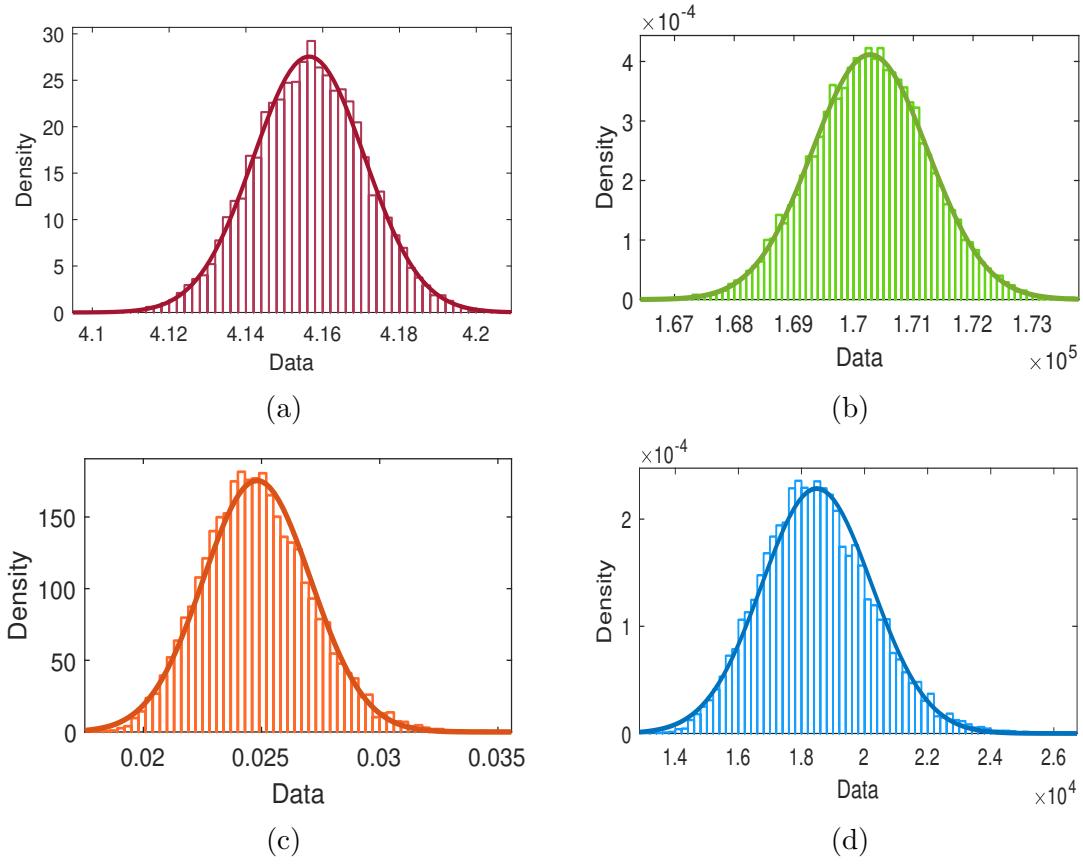


Figure 5.3: After the ensemble is run through the model for a one-year simulation, histograms of each cell population are plotted. (a) cycling stem cells y_1, (b) precursors y_2, (c) mature effector T cells T_1, and (d) regulatory cells R. Each population is fit with a normal distribution. The T_1 and R components of the ensemble are slightly skewed.

used for the forecast model are given by Tables 3.1 and 3.2. The data is described in Section 3.3.1. Performance of the EKF is evaluated by comparison of the data with the estimated observations calculated from the state variable, evolution of the state variables with and without the EKF, evolution of $tr(\boldsymbol{P})$ signifying the mean square error, and evolution of the variance of the state variable given by $\boldsymbol{P}_{i,i}$.

Figure 5.4 shows the estimated BCR-ABL ratio derived with and without the EKF for four representative patients. Initial conditions $\boldsymbol{x_0}$ and $\boldsymbol{P_0}$ are identical for each approach. The pure forecast curves (blue) are identical to the simulations shown in Chapter 3. Assimilating the data using the EKF has varying results. For many patients, the result is quite similar to the pure forecast (e.g. fig. 5.4a). In other cases, the assimilated ratio is pulled towards the data at specific time points. For example, in Figure 5.4d at month 3, the prediction with EKF deviates from, but eventually realigns with the pure forecast. The differences in performance can be explained as follows. First, the forecast model utilizes different parameter values for each patient which change the evolution of both the state and the covariance matrix. Second, the measurement covariance $\boldsymbol{R}$ is a function of the observation. Higher BCR-ABL ratios are associated with a lower CV than low BCR-ABL measurements. Therefore patients who respond better to therapy may have higher variability in their measurements, causing the EKF to correspond more to the model than the data.

Similarity between the pure forecast and the forecast with EKF suggests that the model forecast is weighted more than the observations. This could simply be a reflection of the chosen model error estimates, which if increased could move the assimilated BCR-ABL ratio away from the pure forecast and closer to the data points.

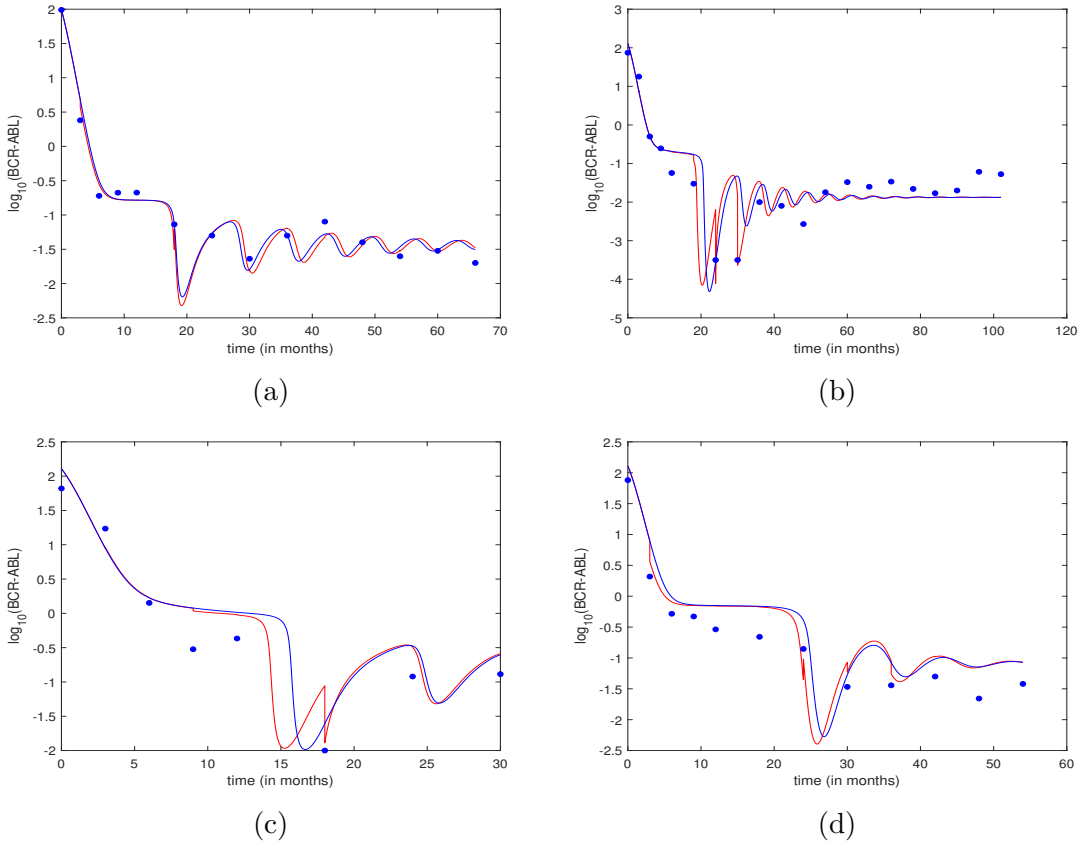


Figure 5.4: Evolution of the BCR-ABL ratio when data is assimilated with EKF for four different patients with $\boldsymbol{Q} = 0$. (a) patient 1, (b) patient 6, (c) patient 8, (d) patient 9. Parameter values are given in tables 3.1 and 3.2. Blue line - forecast only, without assimilation. Red line - with assimilation by EKF. Blue dots - patient data.

Experimentation with inflating certain components of $\boldsymbol{P_0}$ did confirm this for some patients. However, a larger initial covariance also tended to lead to negative, and therefore unrealistic, values of the state variables and variance. In these instances, resetting any negative values to 0 after the update step can, but not always, allow the filter to continue without failing.

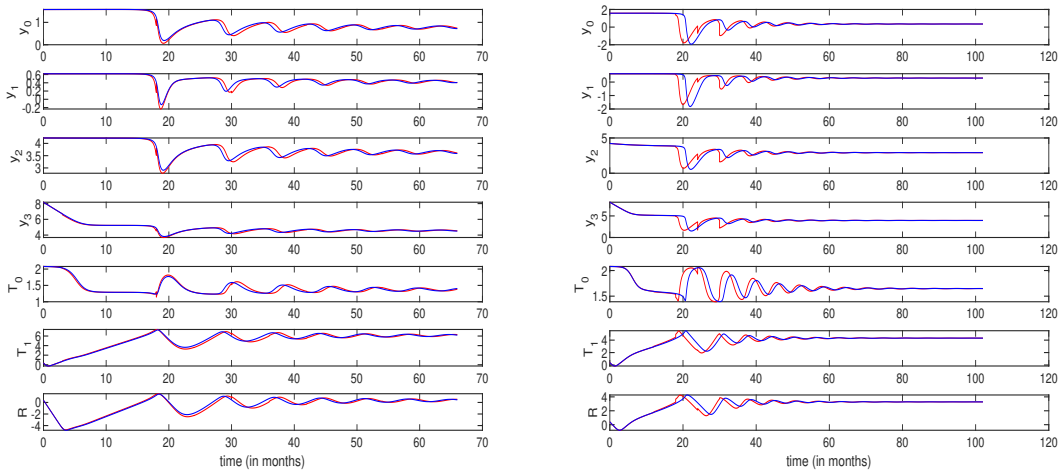
Figure 5.5 shows the evolution of each state variable for the selected patients. As with the BCR-ABL ratios, there is much agreement between the pure forecast and

the forecast with assimilation. Although the observation function $\boldsymbol{h}$ is dependent only on the mature leukemic concentration y_3, the effect of assimilation is similar across all state variables. This suggests a high correlation between these values.

The evolution of the mean square error given by $tr(\boldsymbol{P})$ is depicted in Figure 5.6. For most patients, the evolution of this value is quite similar between the pure forecast and the forecast with EKF. $tr(\boldsymbol{P})$ initially decreases steadily over the first 12 months, after which it increases sometimes almost to its initial value. This increase corresponds to the end of the second decline in the BCR-ABL ratio. A non-monotonic decline occurs for the duration of the assimilation window corresponding to the oscillations in BCR-ABL ratio. Figure 5.7 shows that the evolution of variance for each state variable is quite similar with the exception of T_1 which tends to increase much sooner, during the initial 6 months.

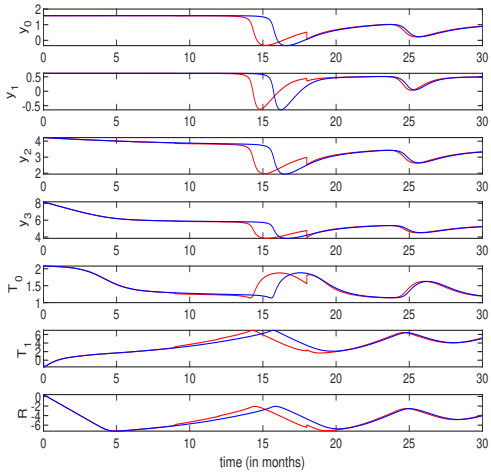
5.3.2 Numerical Results with Process Noise

It is more likely that the model prediction is affected by process noise. In this section, the process noise covariance $\boldsymbol{Q}$ is incorporated through the previously described adaptive method. Various initial conditions were considered for $\boldsymbol{Q}$ as well as for the forgetting factor γ. However, the selection was limited again by a tendency of the state variable to become negative. In certain cases resetting the negative values to 0 allowed the EKF to continue without failing. The choice of $\boldsymbol{Q_0}$ and γ were selected in part by comparing the values of $\boldsymbol{H P^f H^T}$ and $\boldsymbol{R}$ at each update step to retain learning potential from the observations but without causing

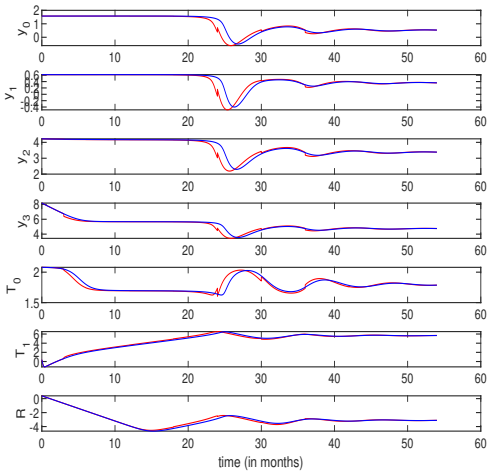


(a)

(b)



(c)



(d)

Figure 5.5: Evolution of each component of the state variable for four different patients. Top to bottom: y_0, y_1, y_2, y_3, T_0, T_1, R. EKF assimilation with $\boldsymbol{Q} = 0$. (a) patient 1, (b) patient 6, (c) patient 8, (d) patient 9. Parameter values can be found in tables 3.1 and 3.2. Blue line - forecast only, without assimilation. Red line - with assimilation by EKF.

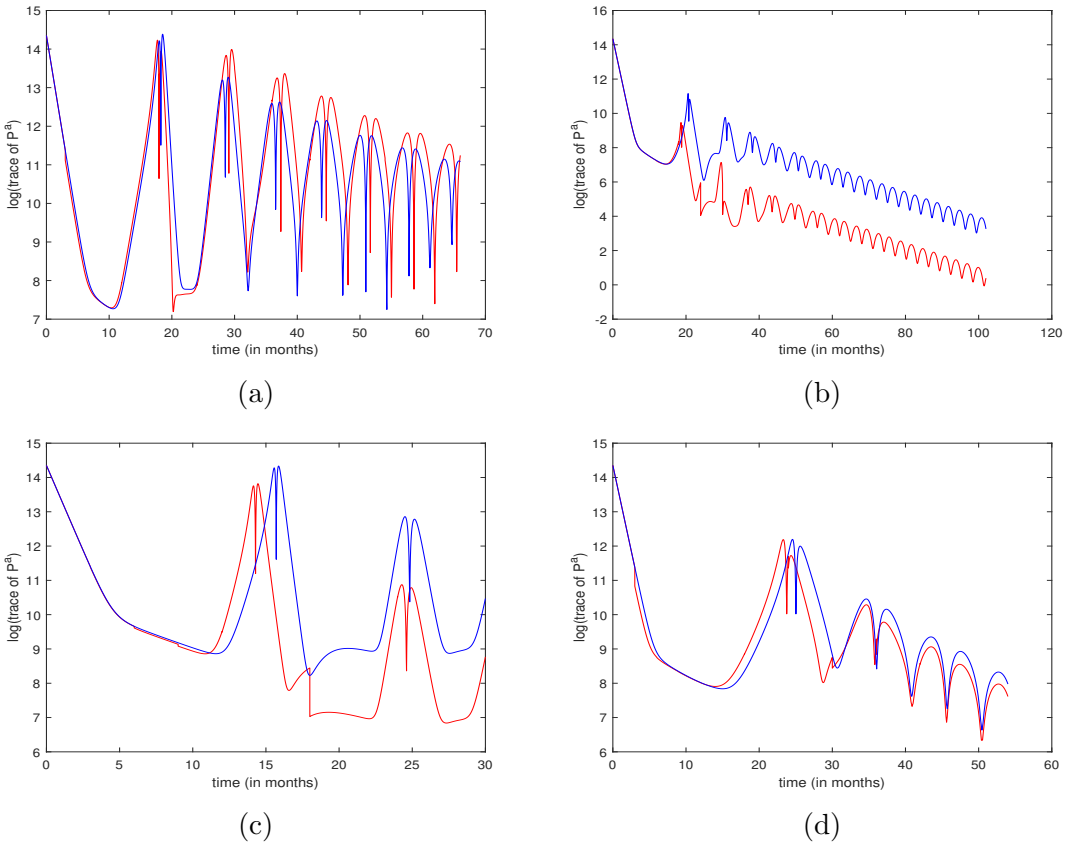


Figure 5.6: The log of $tr(\boldsymbol{P})$ over time for four different patients. EKF assimilation with $\boldsymbol{Q} = 0$. (a) patient 1, (b) patient 6, (c) patient 8, (d) patient 9. Blue line - forecast only, without assimilation. Red line - with EKF assimilation.

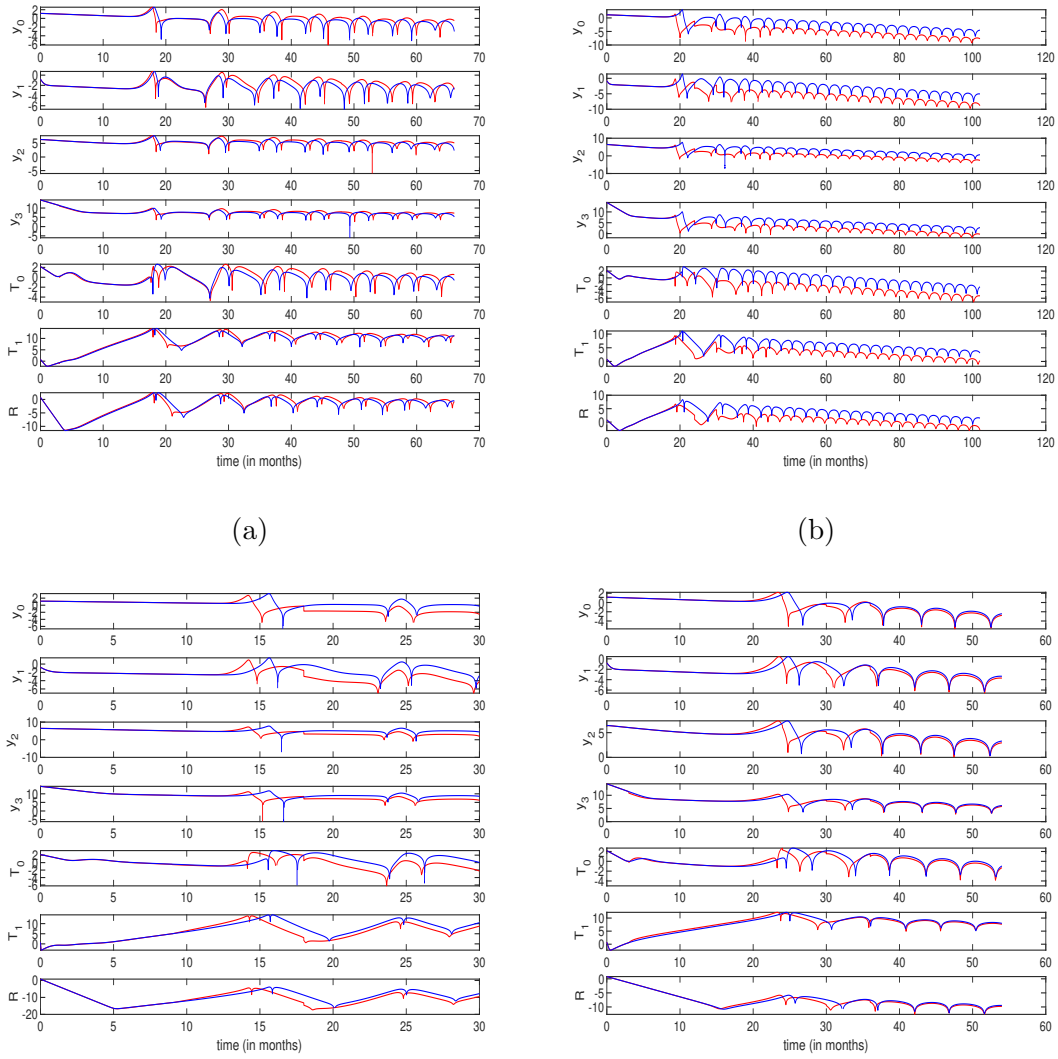


Figure 5.7: The log of variance for each state variable component over time for four different patients. Top to bottom: y_0, y_1, y_2, y_3, T_0, T_1, R. EKF assimilation with $\boldsymbol{Q} = 0$. (a) patient 1, (b) patient 6, (c) patient 8, (d) patient 9. Blue line - forecast only, without assimilation. Red line - with EKF assimilation.

the filter to diverge. An initial process noise covariance of $\boldsymbol{Q_0} = 0$ and a forgetting factor of almost 1 allowed the EKF to perform largely without incident of unrealistic values. This choice puts more emphasis on the original model error covariance and delays the changes due to the innovation. It is not assumed that this choice results in the best performance of the EKF in terms of leading to an accurate prediction.

Figure 5.8 is similar to Figure 5.4 but now incorporates process noise. The evolution of the BCR-ABL ratio is shown for the same four representative patients. The difference in prediction with EKF (red) is most notable towards the end of the assimilation window, as is to be expected with the choice of a large α value. In a sense, $\boldsymbol{Q}$ is acting as an inflation of the error covariance and for many patients pulls the predicted BCR-ABL ratio towards the data over the pure forecast estimation.

The evolution of $tr(\boldsymbol{P})$ is shown in Figure 5.9. The evolution over the first year is similar to the behavior of the same quantity when process noise is not considered. Behavior after that point varies by patient, although it has a tendency to reach a steady state rather than continue to decrease. These two observations are the result of a large α value that delays the inflation of the model covariance.

5.4 Observing System Simulation Experiments

Observing System Simulation Experiments (OSSE) [52] provide a way to design data assimilation ideas and investigate the impact of different types of observations or different frequencies of observations with out the need for real data. An OSSE begins with what is known as a nature run generated from a reasonable model

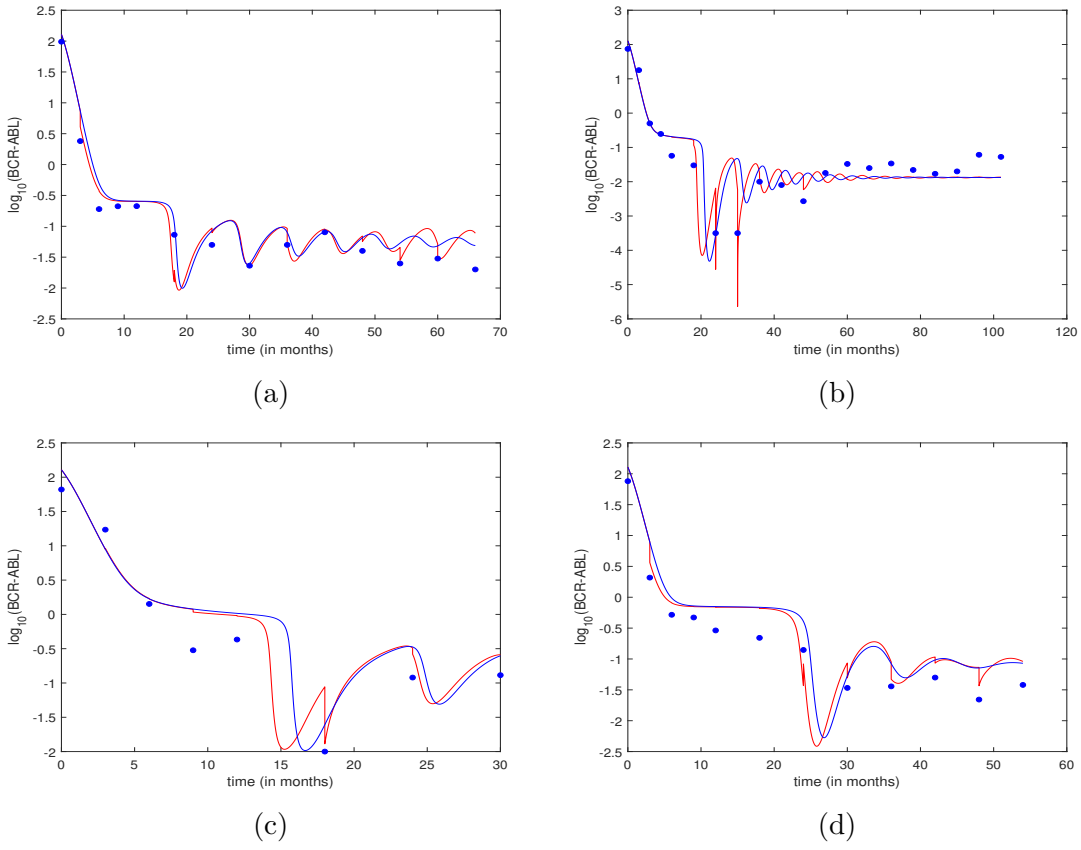


Figure 5.8: Evolution of the BCR-ABL ratio when data is assimilated with EKF for four different patients with adaptive $\boldsymbol{Q}$. (a) patient 1, (b) patient 6, (c) patient 8, (d) patient 9. Parameter values can be found in tables 3.1 and 3.2. Blue line - forecast only, without assimilation. Red line - with assimilation by EKF. Blue dots - patient data.

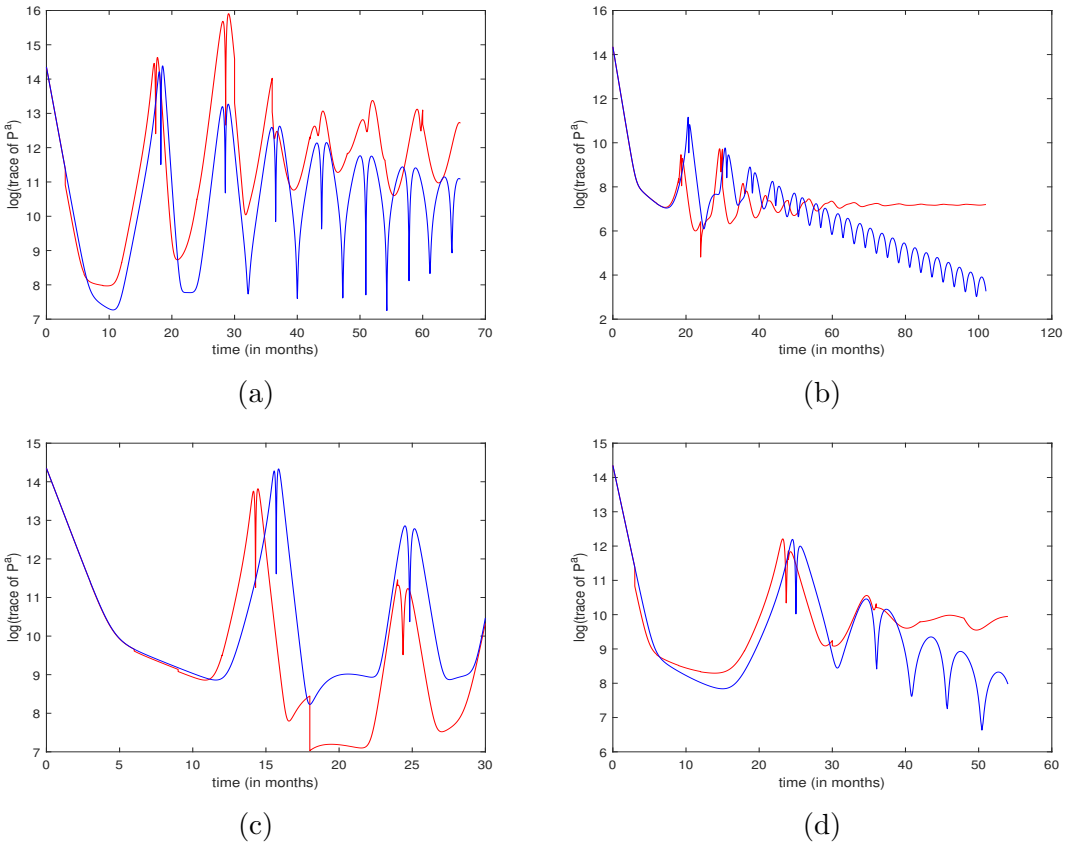


Figure 5.9: The log of $tr(\boldsymbol{P})$ over time for four different patients. EKF assimilation with adaptive $\boldsymbol{Q}$. (a) patient 1, (b) patient 6, (c) patient 8, (d) patient 9. Blue line - forecast only, without assimilation. Red line - with EKF assimilation.

of the system in question. Nature runs ideally have the same statistical behavior as the system that is being model and therefore should be created with realistic, high-quality models. They can be produced by the same dynamical model used in the data assimilation technique, in which case the OSSE is known as an identical-twin experiment. The nature run provides the 'true' state of the system. Synthetic data is then generated from the truth state using the observation function and assigning realistic measurement errors

$$\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}^{\boldsymbol{t}}) + \boldsymbol{\epsilon_m}, \tag{5.9a}$$

$$\boldsymbol{\epsilon_m} \sim \mathcal{N}(0, \boldsymbol{R}^{\boldsymbol{t}}). \tag{5.9b}$$

A control run is performed in which synthetic data representing current available observations is used for the data assimilation scheme. In the context of CML and the extended model (3.1), the control run would consist of synthetic BCR-ABL ratio data measured every 3 months for the first year and every 6 months thereafter. Following the control run, simulated candidate observations are added and assimilated for a perturbation run. Here, two perturbations will be considered. First, we will consider the effect of having more frequent measurements by generating observations that occur every 1 and 3 months throughout treatment. Additionally, new observations in the form of immune cell counts will be considered.

5.4.1 Increased Frequency of Observations

To determine the impact of an increased frequency of observations on the EKF, multiple nature runs were created using both the extended model (3.1) and the Clapp model (2.6), for various patient-specific parameter sets. When using the extended model, only the initial state variable is changed between nature run and the data assimilation process. This identical-twin experiment allows us to test the success of the EKF in recovering the true dynamics of the state. Since the extended model is imperfect, the Clapp model is used for additional experiments to determine how well the EKF can estimate dynamics generated by an alternate model. This imparts a form of model error in the assimilation process.

The following figures show results for three OSSEs exploring the effect of more frequently available data. The nature run for Figure 5.10 is created with the extended model and patient-specific parameter values associated with patient 1 in Table 3.2. Figures 5.11 and 5.12 display results where the true state is generated from the Clapp model with optimal parameters for patients 1 and 6 respectively. For all three experiments, it is clear that the additional observations improve the success of the EKF. Evolution of the BCR-ABL ratio with and without the EKF are quite similar through the first year of treatment. From that point forward, the EKF gradually pulls the ratio closer to the true value. The effect of assimilation increases as the length of time between observations decreases, although the true state is never perfectly recovered. Tuning of the process noise covariance matrix $\boldsymbol{Q}$ may be necessary to achieve this but, as discussed in section 5.3, is difficult to carry

out.

5.4.2 Addition of Immune Assay Data

Since the Clapp model does not differentiate between different types of immune cells, it is difficult to use this model for generating any additional data types. Therefore, only identical-twin experiments were conducted to determine the effect of incorporating immune assay data. Nature runs are generated using the extended model of CML and the immune system with a selected initial condition and set of patient-specific parameters. The same parameters are then used for the model during the forecasting step of the EKF. Observations of the true BCR-ABL ratio and cellular concentrations of T_0, T_1 and R are made every 3 months for the first year and every 6 months thereafter. All four observation types are used to assimilate the model using the EKF. The result is compared to the control run in which only the BCR-ABL ratio is used. The experiment is repeated for various initial state variable values and different patient-specific parameters. The results for one such experiment are shown in Figure 5.13.

It is clear that the addition of new data types greatly improves the ability of the EKF to recover the true state. After only 9 months, the BCR-ABL ratio prediction with EKF using all data types (red) matches the true ratio almost exactly, while the control (green) more closely resembles the pure-forecast ratio (blue). As the assimilation process continues, the perturbation run does appear to drift from the true values but only marginally. It is still much closer to the true value than

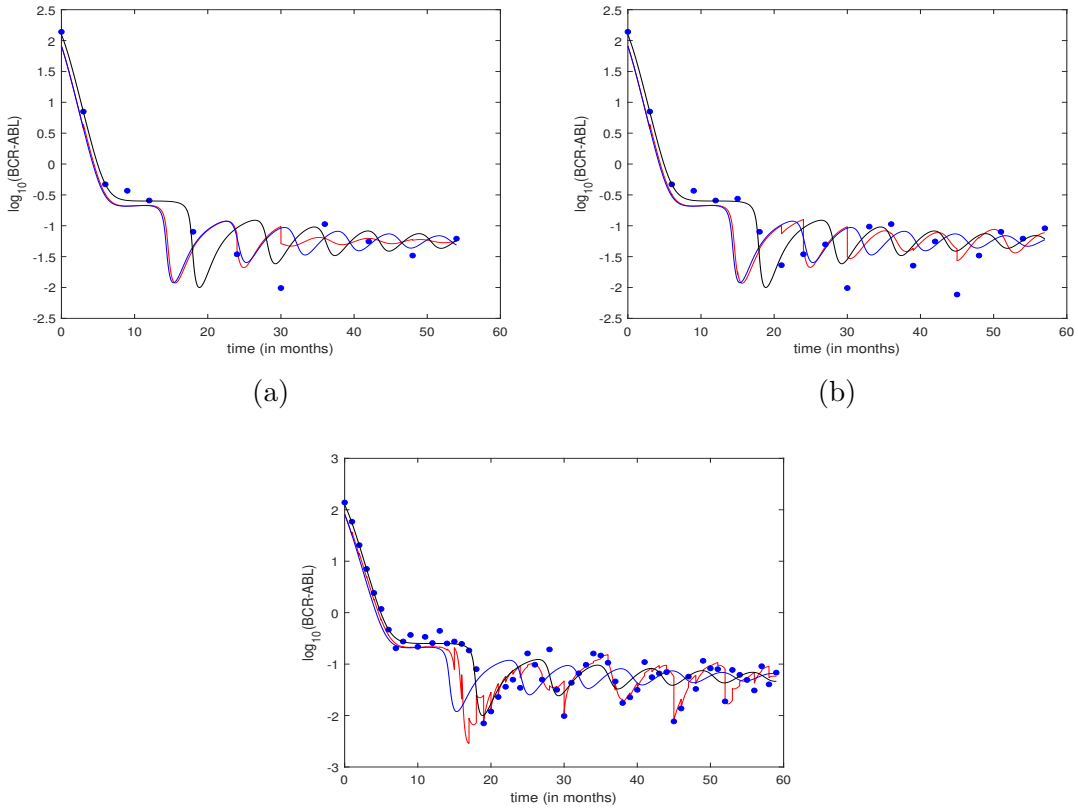


Figure 5.10: Evolution of the BCR-ABL ratio after assimilation with synthetic data. Synthetic data are sampled from a truth run created with the extended model of CML and the immune system ((3.1)). (a) control run, (b) observations every 3 months, (c) observations every month. Black line - true BCR-ABL ratio. Blue line - BCR-ABL ratio without assimilation. Red line - BCR-ABL ratio with EKF assimilation. Blue dots - noisy observations of the true state.

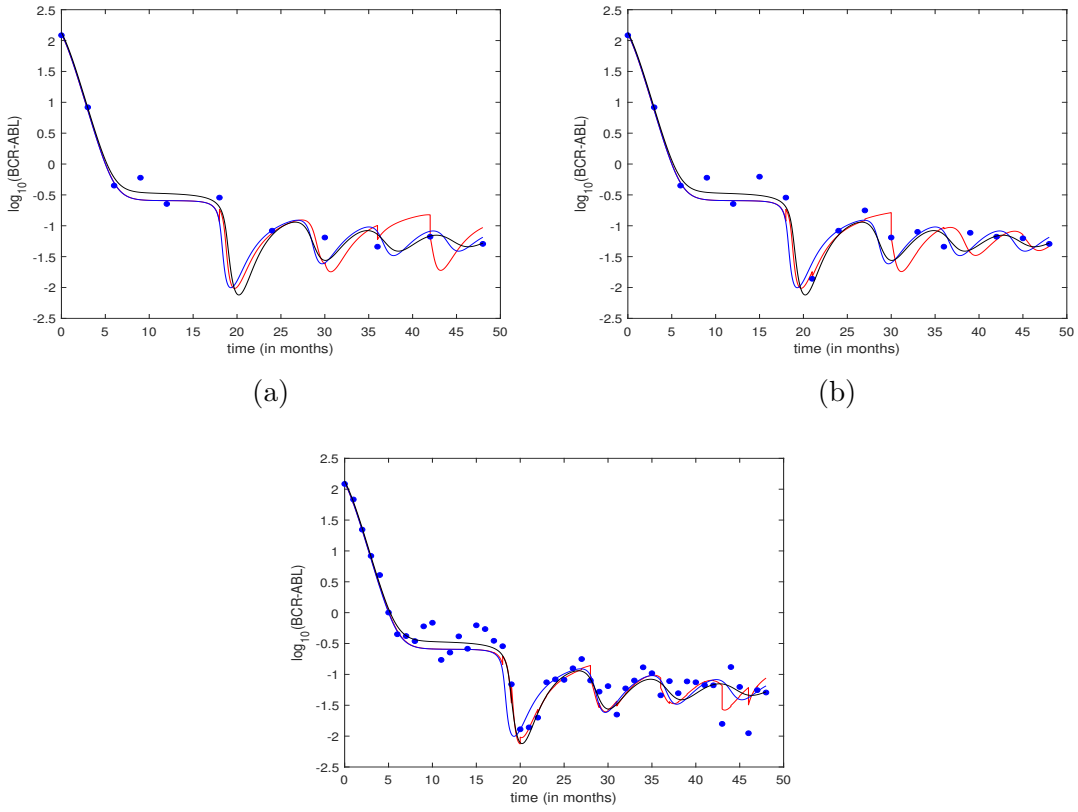
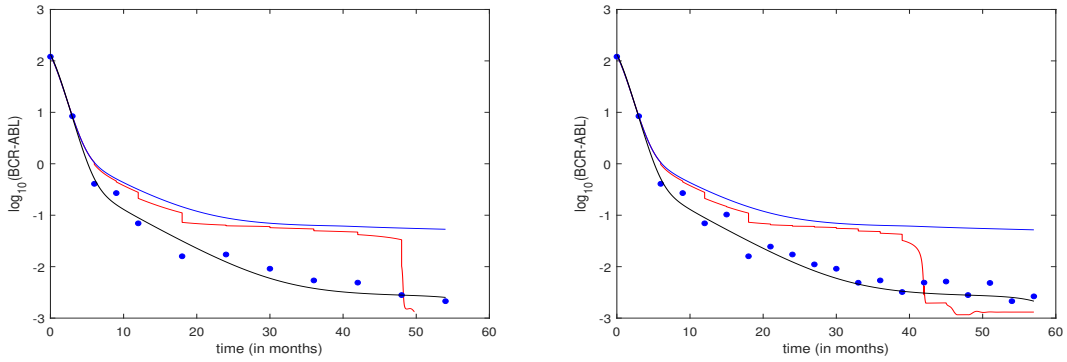
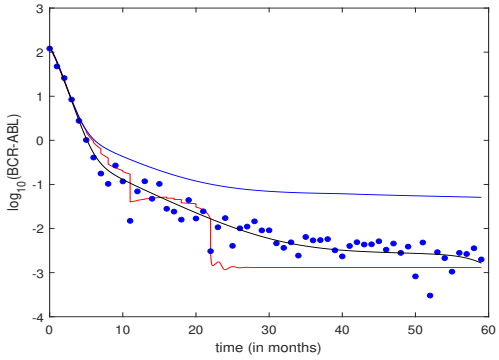


Figure 5.11: Evolution of the BCR-ABL ratio after assimilation with synthetic data. Synthetic data are sampled from a truth run created with the Clapp model ((2.6)). (a) control run, (b) observations every 3 months, (c) observations every month.Black line - true BCR-ABL ratio. Blue line - BCR-ABL ratio without assimilation. Red line - BCR-ABL ratio with EKF assimilation. Blue dots - noisy observations of the true state.



(a) (b)



(c)

Figure 5.12: Evolution of the BCR-ABL ratio after assimilation with synthetic data. Synthetic data are sampled from a truth run created with the Clapp model ((2.6)). Patient-specific parameter values differ from those used in fig. 5.11. (a) control run, (b) observations every 3 months, (c) observations every month. Black line - true BCR-ABL ratio. Blue line - BCR-ABL ratio without assimilation. Red line - BCR-ABL ratio with EKF assimilation. Blue dots - noisy observations of the true state.

the control run. The same behavior is demonstrated in the evolution of the three immune cell populations.

The effect of incorporating these new data types is immediate and lasting. This would suggest that immune assay data does not need to be measured as frequently as the BCR-ABL ratio for the EKF to be successful. Observations should be taken through the first year of treatment and near any changes in therapy that could likely result in a change in disease dynamics (dose or scheduling changes, cessation, etc.).

5.5 Discussion

Data assimilation has proven to be an extremely useful process for combining imperfect dynamical models of a process with noisy observations to produce reasonably likely estimates of the true state. While the use of data assimilation in weather forecasting and other geoscientific applications is well established, it has only recently been considered as a useful tool in biomedical applications. The ability to accurately predict disease progression would allow doctors to proactively tune treatment protocols for individual patients, and therefore improve outcomes. With respect to CML, data assimilation could aid in the creation of more predictive models that can be used to predetermine relapse in treatment cessation trials.

Since the extended model of CML and the immune system is nonlinear and the assumption of gaussian model noise seemed reasonable, the extended Kalman filter was selected as an appropriate data assimilation technique. Effect of the EKF

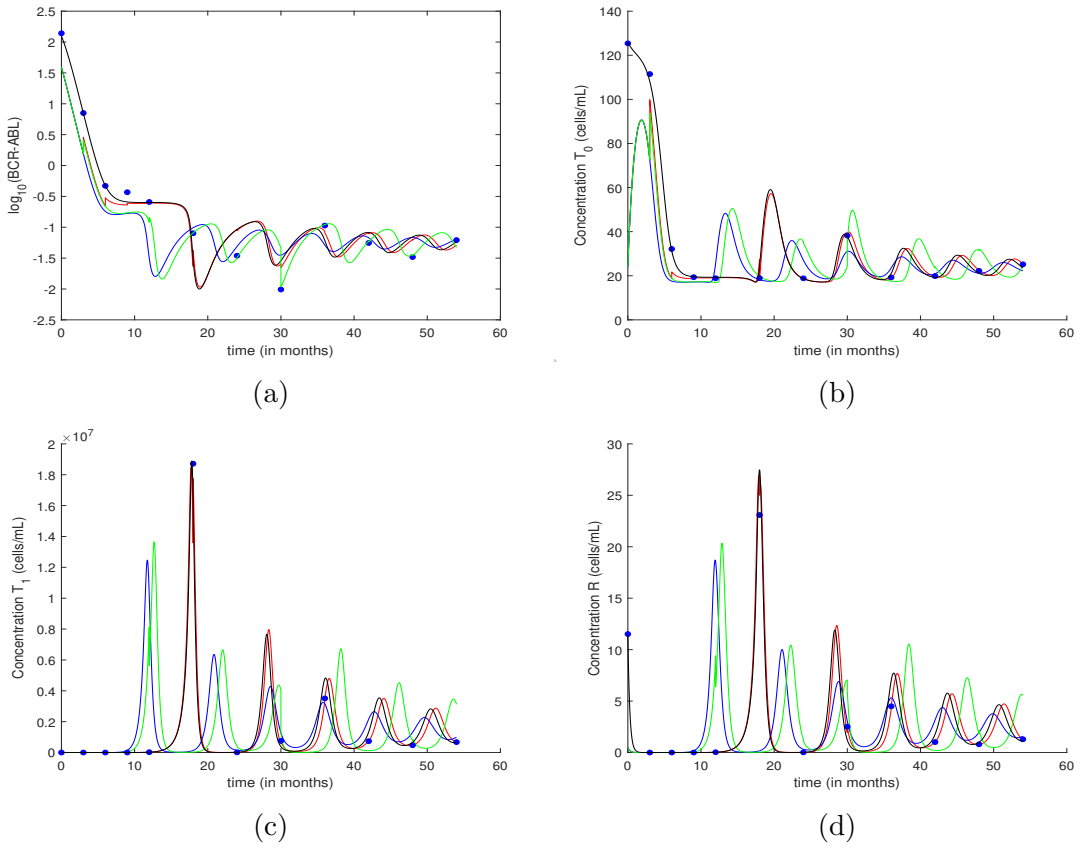


Figure 5.13: Evolution of the BCR-ABL ratio after assimilation with synthetic data. Synthetic data are sampled from a truth run created with the Clapp model ((2.6)). Patient-specific parameter values differ from those used in fig. 5.11. (a) log of the BCR-ABL ratio, (b) cell concentration of T_0, (c) cell concentration of T_1, (d) cell concentration of R. Black line - true value. Blue line - forecast only, without assimilation. Red line - with assimilation by EKF using all 4 data types. Green line - control run, only BCR-ABL ratio used in EKF. Blue dots - noisy observations of the true state.

was shown to vary patient to patient most likely due to the high degree of variation in patient-specific parameter values. For some patients, the EKF produced very little change from the pure prediction of the model, while for others the forecast was noticeably different.

Success of the data assimilation process is difficult to determine from the available patient data alone, since the true state itself is unknown. To better evaluate performance, a variety of OSSEs were performed in which the ability of the EKF process to recover a 'true' state was tested using synthetic data. We found that the EKF is able to adjust the state variable to a point that more closely resembles the true state, although it was not able to consistently recover the exact true state. It was determined that more frequent observations resulted in a more accurate recovery. Based on these results, we could recommend to increase the frequency of data collection to at least every 3 months throughout treatment. Measurements every month would be ideal, but realistically may be inconvenient and expensive.

Additionally, incorporation of more data types can greatly improve results. By including three new data types that directly measure the concentrations of immune cells, the forecast recovers the true state almost immediately. It does not appear that these cell concentrations need to be measured as frequently as the BCR-ABL ratio alone. Obtaining this data for the first year of therapy appears to be enough to successfully recover the underlying dynamics, assuming accurate parameter values.

The success of any data assimilation processes is heavily weighted on the ability to accurately estimate the covariances of process noise $\boldsymbol{Q}$ and measurement error $\boldsymbol{R}$. We consider the measurement errors used in this work to be reasonable. However,

it may be possible to obtain more precise estimates directly from our collaborators. The appropriate estimate for $\boldsymbol{Q}$ was more difficult to acquire. While the adaptive process used in this chapter has been shown to be more robust to initial errors than other methods [43], tuning the initial guess and the forgetting factor α proved difficult due to the tendency of the updated state variable values and variances to become negative. This is perhaps due to the possibly incorrect assumption of Gaussianity in the model error. Further investigation is needed or perhaps an alternate data assimilation technique. For example, a particle filter could be utilized, which makes no assumption about specific error distributions.

Error in parameter values, particularly patient-specific parameters, was not taken into account in this work. Since inaccurate values could affect the EKF process, these errors should be addressed before results of the EKF on patient data can be fully evaluated. Data assimilation techniques can be extended for optimal parameter estimation by simply augmenting the state variable with the parameters of interest.

$$\boldsymbol{x} = \begin{bmatrix} x_s \\ \\ x_p \end{bmatrix}. \tag{5.10}$$

The new state variable consists of two components, $\boldsymbol{x_s}$ the state variable as described in Sections 5.2 and 5.3 and $\boldsymbol{x_p}$ representing the model parameters. $\boldsymbol{x}$ is an N-

dimensional vector where $N = N_s + N_p$ and evolves according to

$$\frac{d\boldsymbol{x}}{dt} = \frac{d}{dt}\begin{bmatrix}\boldsymbol{x_s}\\ \\ \boldsymbol{x_p}\end{bmatrix} = \begin{bmatrix}\boldsymbol{f}(\boldsymbol{x_s})\\ \\ 0\end{bmatrix}. \tag{5.11}$$

This assumes that the model parameters are subject only to stochastic processes. The new model error covariance matrix is defined as

$$\boldsymbol{P} = \begin{bmatrix}\boldsymbol{P_{ss}} & \boldsymbol{P_{sp}}\\ \\ \boldsymbol{P}_{ps} & \boldsymbol{P}_{pp}\end{bmatrix} \equiv \mathbb{E}\left(\begin{bmatrix}\boldsymbol{x_s} - \boldsymbol{x_s^t}\\ \\ \boldsymbol{x_p} - \boldsymbol{x_p^t}\end{bmatrix}\begin{bmatrix}\boldsymbol{x_s} - \boldsymbol{x_s^t}\\ \\ \boldsymbol{x_p} - \boldsymbol{x_p^t}\end{bmatrix}^T\right) \tag{5.12}$$

Thus $\boldsymbol{P}$ is an $(N_s + N_p) \times (N_s + N_p)$ matrix. Likewise, $\boldsymbol{Q}$ accounts for errors associated with both the system components and the model parameters. In the context of our CML model (3.1), $\boldsymbol{x_p}$ would be an 8-dimensional vector consisting of the patient-specific parameter values inh_1, inh_2, d_{T1}, μ, y_{min}, y_{max}, k_T and r_T. Since these values are patient-specific, the covariance $\boldsymbol{P_{pp}}$ is likely also dependent on the specific patient and could potentially be determined using the adaptive parameter values computed in Chapter 4. Besides changes in dimensionality and updates of the TLM $\boldsymbol{F}$ and Jacobian of the observation function $\boldsymbol{H}$, the forecast and analysis steps of the EKF would remain as previously stated in Section 5.2.

Although the use of data assimilation techniques in modeling CML needs to be explored further, the results described in this chapter suggest its use to improve the predictive power of CML models appears promising.

Chapter 6: Conclusion

Since the development of TKIs in the early 2000s, CML has become a manageable chronic condition increasing survival rate and leading approximately 80% of patients to achieve CCR [3]. Despite this, TKIs are not considered a cure. Most patients will take the drug indefinitely to prevent progression to the highly lethal blast phase. This can be both inconvenient and costly, affecting patient quality of life. Thus, much of the clinical research surrounding CML is conducted in an effort to increase the number of patients reaching TFR and determining characteristics necessary to maintain TFR.

Research into both of these areas suggests involvement of the immune system. Prior to the development of TKIs, IFNα was considered the primary treatment for CML. Discovery of numerous anti-tumor and immunomodulatory effects, have led to creation of treatment protocols combining IFNα with TKIs. In the last decade, outcomes of treatment cessation trials such as STIM [14], TWISTER [15] and Euro-Ski [16] have further increased the interest in understanding the role of immune response in CML. Although different criteria for candidate patients and different definitions of relapse were utilized across these trials, the outcome of each determined that successful treatment cessation is possible. Between 40% and 50%

of patients experienced TFR following treatment cessation, often with low levels of leukemic cells still present. The immune system is hypthesized to keep this residual disease in check, thus preventing a relapse. Determining the exact involvement of the immune system and the appropriate criteria for treatment cessation candidates remains an open question in clinical research.

Mathematical modeling provides a beneficial complementing approach to clinical research, allowing for the exploration of the underlying dynamics of the disease and experimentation with treatment protocols in a cost-effective manner. Various models of leukemia, and CML in particular, have been created. Early models of CML under TKI therapy focus on capturing the initial biphasic decline in patient BCR-ABL ratios, providing explanation for the phenomenon. Findings in clinical trials and research tying immune response to disease management inspired new mathematical models incorporating the immune system in some capacity. The Kim [21] and Clapp [22] models incorporated the autologous immune response through a single immune cell compartment. This addition allowed Clapp *et al.* to provide explanation for the oscillations often observed in patient data in later phases of TKI treatment. While their results provide new insights into the underlying biological processes, the Clapp model is unable to explain extended relapse-free periods following treatment cessation and cannot accurately forecast disease progression. The goal of the research presented in this dissertation was to develop new mathematical models and consider numerical methods that together could be utilized to provide reasonable predictions of CML for use in a clinical setting.

A new model of CML and the immune system is presented in Chapter 3, which

is an extension of the Clapp model [22]. The Clapp model describes CML through 5 interacting compartments, 4 of which represent leukemic cells at various stages of maturation and the last representing a generic immune cell. The dynamics governing leukemic cells have been well established in numerous mathematical models. For this reason, these equations were left largely untouched. Instead, the focus of our model was to provide a more detailed description of the immune response by expanding the single immune compartment of the Clapp model (z) into three new compartments (T_0, T_1, R) incorporating adaptive regulation of immune response through the addition of regulatory T cells.

The extended model (3.1) was fit to patient data and compared to the performance of the Clapp model. Although the two models do not produce identical simulations, they both perform qualitatively well in that they capture all of the key characteristics of CML patient data. The extended model may be better suited for determining the difference between relapse and remission. For example, it was shown that reducing the concentration of regulatory cells can greatly increase the likelihood of TFR. Nonetheless, the extended model does not have much predictive power. When simulating treatment cessation by setting TKI inhibition parameters to 1, the model consistently predicts an immediate relapse. Adjusting the immune parameters can allow for the simulation of remission, but it is unclear how this should be done. It is evident that in order to use the extended model for prediction purposes, the model parameters must be better identified.

In Chapter 4, we considered the possibility that the patient-specific parameters could vary in time and developed a routine to refit the model to the data in different

time windows. The routine begins by fitting the model only to data that appears in the first window. A window length of two years allowed major characteristics of the data to be captured in a single window and worked well with the frequency of our data. Since biologically the behavior of the immune system is not considered to drastically change over time, a second penalty was added to the objective function in the optimization scheme to prevent large changes in parameter values from window to window. The results suggest that the parameter values governing the effect of therapy and the size of the immune window are well established and are unlikely to change over the course of treatment. However, the remaining patient-specific parameters fluctuate much more across the windows. Accurately estimating these values will be key to fully understanding the role of immune response in CML.

The adaptive parameter fitting routine was used to fit the model to Stop TKI data with the goal of discovering characteristics that could be used to distinguish between relapse and TFR patients. We were unable to find such characteristics that clearly categorized the two potential results of treatment cessation. However, it was noted that immune parameters of relapse patients adapted in a way that reflected a diminished immune response in later stages of therapy.

Data assimilation techniques have long been used in the geosciences to improve understanding of underlying dynamics and improve predictions of physical processes. Application of data assimilation in biomedicine is relatively new, yet holds promising prospects for creating mathematical models that can be used for prediction in a clinical setting. In Chapter 5, the extended Kalman filter was used to assimilate our extended model of CML with patient data. The performance of the EKF using real

patient data was not well-determined as the update step had a tendency to produce unrealistic values for our state variable. However, OSSEs suggested that the EKF could be successful if parameter values are well established. Results of various OSSEs also suggested that an increased frequency of measurements as well as availability of new data types in the form of immune assays could be vital to success of the assimilation process. Data assimilation can also be used for parameter estimation. This was not attempted in the course of the research presented here, but should be considered for future work.

Collaboration between the mathematical and medical communities can provide insight to disease in a way that neither can provide on their own. Data provided by clinicians are used to inform mathematical models, which in turn can be used to address questions that are difficult to answer in a clinical setting. Continued cooperation between these communities is vital in the search for a cure to cancer, particularly in the context of CML where many questions about treatment cessation remain unanswered.

Bibliography

[1] Michele Cea, Antonia Cagnetta, Alessio Nencioni, Marco Gobbi, and Franco Patrone. New insights into biology of chronic myeloid leukemia: Implications in therapy. *Current Cancer Drug Targets*, 13(7):711–723, 2013.

[2] Alfonso Quintás-Cardama and Jorge E. Cortes. Chronic myeloid leukemia: Diagnosis and treatment. *Mayo Clinic proceedings*, 81(7):973–88, 07 2006. Copyright - Copyright Mayo Foundation for Medical Education and Research Jul 2006; Last updated - 2017-11-09; CODEN - MACPAJ.

[3] National Cancer Institute. Chronic myelogenous leukemia treatment, 2016.

[4] Bradley Chereda and Junia V. Melo. Natural course and biology of cml. *Annals of Hematology*, 94(2):107–121, Apr 2015.

[5] Razelle Kurzrock, Hagop M. Kantarjian, Brian J. Druker, and Moshe Talpaz. Philadelphia ChromosomePositive Leukemias: From Basic Mechanisms to Molecular Therapeutics. *Annals of Internal Medicine*, 138(10):819–830, 05 2003.

[6] Xin An, Amit K. Tiwari, Yibo Sun, Pei-Rong Ding, Charles R. Ashby, and Zhe-Sheng Chen. Bcr-abl tyrosine kinase inhibitors in the treatment of philadelphia chromosome positive chronic myeloid leukemia: A review. *Leukemia Research*, 34(10):1255 – 1268, 2010.

[7] American Cancer Society. Cancer facts and figures 2016, 2016.

[8] Branford, Hughes, and Rudzki. Monitoring chronic myeloid leukaemia therapy by real-time quantitative pcr in blood is a reliable alternative to bone marrow cytogenetics. *British Journal of Haematology*, 107(3):587–599, 1999.

[9] Susan Branford, Linda Fletcher, Nicholas C. P. Cross, Martin C. Müller, Andreas Hochhaus, Dong-Wook Kim, Jerald P. Radich, Giuseppe Saglio, Fabrizio Pane, Suzanne Kamel-Reid, Y. Lynn Wang, Richard D. Press, Kevin Lynch,

Zbigniew Rudzki, John M. Goldman, and Timothy Hughes. Desirable performance characteristics for bcr-abl measurement on an international reporting scale to allow consistent interpretation of individual patient response and comparison of response rates between clinical trials. *Blood*, 112(8):3330–3338, 2008.

[10] D. Ribatti. The concept of immune surveillance against tumors. the first theories. *Oncotarget*, 8(4):7175–7180, 2016.

[11] Peter Rohon. Biological therapy and the immune system in patients with chronic myeloid leukemia. *International Journal of Hematology*, 96(1):1–9, Jul 2012.

[12] Anna Kreutzman, Vesa Juvonen, Veli Kairisto, Marja Ekblom, Leif Stenke, Ruth Seggewiss, Kimmo Porkka, and Satu Mustjoki. Mono/oligoclonal t and nk cells are common in chronic myeloid leukemia patients at diagnosis and expand during dasatinib therapy. *Blood*, 116(5):772–782, 2010.

[13] M. Talpaz, J. Mercer, and R. Hehlmann. The interferon-alpha revival in cml. *Annals of Hematology*, 94(2):195–207, 2015.

[14] FX Mahon, D. Rea, J. Guilhot, F. Guilhot, F. Huguet, and et al. F Nicolini. Discontinuation of imatinib in patients with chronic myeloid leukaemia who have maintained complete molecular remission for at least 2 years: the prospective, multicentre stop imatinib (stim) trial. *Lancet Oncology*, 11(11):1029–1035, 2010.

[15] DM Ross, S. Branford, JF Seymour, AP Schwarer, C. Arthur, and et al. DT Yeung. Safety and efficacy of imatinib cessation for cml patients with stable undetectable minimal residual disease: results from the twister study. *Blood*, 122(4):515–522, 2013.

[16] Francois-xavier Mahon, Johan Richter, Joelle Guilhot, Henrik Hjorth-Hansen, Antonio Almeida, Jeroen J.W.M. JWM Janssen, Jiri Mayer, Kimmo Porkka, Panayiotis Panayiotidis, Ulla Stromberg, Marc G Berger, Joanna Diamond, Hans Ehrencrona, Veli Kairisto, Katerina Machova Polakova, Martin C. Mueller, Satu Mustjoki, Andreas Hochhaus, Markus Pfirrmann, and Susanne Saussele. Cessation of tyrosine kinase inhibitors treatment in chronic myeloid leukemia patients with deep molecular response: Results of the euro-ski trial. *Blood*, 128(22):787–787, 2016.

[17] S Saußele, J Richter, A Hochhaus, and F-X Mahon. The concept of treatment-free remission in chronic myeloid leukemia. *Leukemia*, 30(8):1638—1647, August 2016.

[18] G. Clapp and D. Levy. A review of mathematical models for leukemia and lymphoma. *Drug Discovery Today: Disease Models*, 16, 2015.

[19] I. Roeder, M. Horn, I. Glauche, A. Hochhaus, M.C. Mueller, and M. Loeffler. Dynamic modeling of imatinib-treated chronic myeloid leukemia: functional insights and clinical implications. 12(10):1181–1184, 2006.

[20] P.S. Kim, P.P. Lee, and D. Levy. Modeling imatinib-treated chronic myelogeneous leukemia: reducing the complexity of agent-based models. 70(3):1181–1184, 2008.

[21] Peter S Kim, Peter Lee, and Doron Levy. Dynamics and potential impact of the immune response to chronic myelogenous leukemia. *PLoS computational biology*, 4:e1000095, 07 2008.

[22] G. Clapp, T. Lepoutre, R. El Cheikh, S. Bernard, J. Ruby, and H. Labussiere-Wallet. Implication of the autologous immune system in bcr-abl transcript variations in chronic myelogeneous leukemia patients treated with imatinib. *Cancer Research*, 75(19):4053–62, 2015.

[23] Raluca Eftimie, Joseph J. Gillard, and Doreen A. Cantrell. Mathematical models for immunology: Current state of the art and future research directions. *Bulletin of Mathematical Biology*, 78(10):2091–2134, Oct 2016.

[24] Raluca Eftimie, Jonathan L. Bramson, and David J. D. Earn. Interactions between the immune system and cancer: A brief review of non-spatial mathematical models. *Bulletin of Mathematical Biology*, 73(1):2–32, Jan 2011.

[25] Yoram Louzoun. The evolution of mathematical immunology. *Immunological reviews*, 216:9–20, 05 2007.

[26] P. Kim, P. Lee, and D. Levy. Basic principles in modeling adaptive regulation and immunodominance. In et al U. Ledzewicz, editor, *Mathematical Methods and Models in Biomedicisne*, pages 33–57. Springer, New York, 2013.

[27] Richard J. Jones, William H. Matsui, and B. Douglas Smith. Cancer Stem Cells: Are We Missing the Target? *JNCI: Journal of the National Cancer Institute*, 96(8):583–585, 2004.

[28] P. Brodin and MM Davis. Human immune System variation. *Nature reviews. Immunology*, 17(1):21–29, 2017.

[29] L. Sompayrac. *How Cancer Works*. How Cancer Works. Jones and Bartlett Publishers, 2004.

[30] Franziska Michor, Timothy P Hughes, Yoh Iwasa, Susan Branford, Neil P Shah, Charles L Sawyers, and Martin A Nowak. Dynamics of chronic myeloid leukaemia. *Nature*, 435(7046):1267, 2005.

[31] Rob DeConde, Peter S Kim, Doron Levy, and Peter P Lee. Post-transplantation dynamics of the immune response to chronic myelogenous leukemia. *Journal of theoretical biology*, 236(1):39–59, 2005.

[32] Artur César Fassoni, Ingo Roeder, and Ingmar Glauche. To cure or not to cure: Consequences of immunological interactions in cml treatment. *Bulletin of Mathematical Biology*, 81(7):2345–2395, Jul 2019.

[33] E John Wherry. T cell exhaustion. *Nature immunology*, 12(6):492, 2011.

[34] Eric Kostelich, Yang Kuang, Joshua M. McDaniel, Nina Z. Moore, Nikolay L. Martirosyan, and Mark C. Preul. Accurate state estimation from uncertain data and models: An application of data assimilation to mathematical models of human brain tumors. *Biology Direct*, 6, 12 2011.

[35] M. J. Hoffman, N. S. LaVigne, S. T. Scorse, F. H. Fenton, and E. M. Cherry. Reconstructing three-dimensional reentrant cardiac electrical wave dynamics using data assimilation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(1):013107, 2016.

[36] M. Sermesant, P. Moireau, O. Camara, J. Sainte-Marie, R. Andriantsimiavona, R. Cimrman, D.L.G. Hill, D. Chapelle, and R. Razavi. Cardiac function estimation from mri using a heart model and data assimilation: Advances and difficulties. *Medical Image Analysis*, 10(4):642 – 656, 2006. Special Issue on Functional Imaging and Modelling of the Heart (FIMH 2005).

[37] Luís M. A. Bettencourt, Ruy M. Ribeiro, Gerardo Chowell, Timothy Lant, and Carlos Castillo-Chavez. Towards real time epidemiology: Data assimilation, modeling and anomaly detection of health surveillance data streams. In Daniel Zeng, Ivan Gotham, Ken Komatsu, Cecil Lynch, Mark Thurmond, David Madigan, Bill Lober, James Kvach, and Hsinchun Chen, editors, *Intelligence and Security Informatics: Biosurveillance*, pages 79–90, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[38] Franz Hamilton, John Cressman, Nathalia Peixoto, and Timothy Sauer. Reconstructing neural dynamics using data assimilation with multiple models. *EPL (Europhysics Letters)*, 107(6):68005, sep 2014.

[39] TE Yankeelov, V. Quaranta, KJ Evans, and EC Rericha. Toward a science of tumor forecasting for clinical oncology. *Cancer Research*, 75(6):918–923, 2015.

[40] Christopher K. Wikle and L. Mark Berliner. A bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, 230(1):1 – 16, 2007. Data Assimilation.

[41] Kayo Ide and Michael Ghil. Extended kalman filtering for vortex systems. part 1: Methodology and point vortices. *Dynamics of Atmospheres and Oceans*, 27(1-4):301–332, 1998.

[42] Jim Kao, Dawn Flicker, Rudy Henninger, Sarah Frey, Michael Ghil, and Kayo Ide. Data assimilation with an extended kalman filter for impact-produced shock-wave dynamics. *Journal of Computational Physics*, 196(2):705 – 723, 2004.

[43] Shahrokh Akhlaghi, Ning Zhou, and Zhenyu Huang. Adaptive adjustment of noise covariance in kalman filter for dynamic state estimation. In *2017 IEEE Power & Energy Society General Meeting*, pages 1–5. IEEE, 2017.

[44] Amy Hughes, Jade Clarson, Carine Tang, Ljiljana Vidovic, Deborah White, Timothy Hughes, and Agnes Yong. Cml patients with deep molecular responses to tki have restored immune effectors, decreased pd-1 and immune suppressors. *Blood*, 129, 01 2017.

[45] Siti-Zuleha Idris, Norfarazieda Hassan, Le-Jie Lee, Sabariah Md Noor, Raudhawati Osman, Marsitah Abdul-Jalil, Abdul-Jalil Nordin, and Maha Abdullah. Increased regulatory t cells in acute lymphoblastic leukemia patients. *Hematology*, 20(9):523–529, 2015. PMID: 26119924.

[46] Xingbing Wang, Jine Zheng, Jun Liu, Junxia Yao, Yanli He, Xiaoqing Li, Jingming Yu, Jing Yang, Zhongping Liu, and Shiang Huang. Increased population of cd4+cd25high regulatory t cells with their higher apoptotic and proliferating status in peripheral blood of acute myeloid leukemia patients. *European Journal of Haematology*, 75(6):468–476, 2005.

[47] Brent Wood, Dragan Jevremovic, Marie C. Béné, Ming Yan, Patrick Jacobs, Virginia Litwin, and ; on behalf of ICSH/ICCS Working Group. Validation of cell-based fluorescence assays: Practice guidelines from the icsh and iccs – part v – assay performance criteria. *Cytometry Part B: Clinical Cytometry*, 84(5):315–323, 2013.

[48] Mark Shenkin, Ramesh Babu, and Russell Maiese. Accurate assessment of cell count and viability with a flow cytometer. *Cytometry Part B: Clinical Cytometry*, 72B(5):427–432, 2007.

[49] Mohamed Al-Rubeai, Karl Welzenbach, D Lloyd, and A Emery. A rapid method for evaluation of cell number and viability by flow cytometry. *Cytotechnology*, 24:161–8, 07 1997.

[50] Weiyu Yang and Ionel Navon. Documentation of the tangent linear model and its adjoint of the adiabatic version of the nasa geos-1 c-grid gcm (version 5.2). 09 1995.

[51] Jurgen A Doornik and Henrik Hansen. An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70:927–939, 2008.

[52] Michiko Masutani, Thomas W. Schlatter, Ronald M. Errico, Ad Stoffelen, Erik Andersson, William Lahoz, John S. Woollen, G. David Emmitt, Lars-Peter Riishøjgaard, and Stephen J. Lord. *Observing System Simulation Experiments*, pages 647–679. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.